150.51 V.45 B
Q82 54.

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *Human Resources Research Center*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

FEB 24 1953
LIBRARY

CONTENTS

Transfer in Motor Learning As a Function of Degree of First-Task Learning and Inter-Task Similarity: C. P. DUNCAN 1

The Law of Primary Reinforcement in Children: P. S. SIEGEL AND J. G. FOSHEE 12

Assimilation of Information from Dot and Matrix Patterns: E. T. KLEMMER AND F. C. FRICK 15

Drive Conditioning As a Factor in Latent Learning: P. E. MEEHL AND K. MACCORQUODALE.. 20

Number of Common Elements and Consistency of Reinforcement in a Discrimination Learning Task: R. S. FRENCH 25

Stimulus Generalization As a Function of Level of Experimentally Induced Anxiety: G. ROSENBAUM 35

The Role of Instruction in Experimental Space Perception: C. W. MANN AND R. O. BORING 44

Effects of Intense Pure Tone Stimuli When Magnitude of Initial Injury Is Controlled: I. E. ALEXANDER AND F. J. GITHLER 49

Learning and Performance in Eyelid Conditioning As a Function of Intensity of the UCS: K. W. SPENCE 57

Perception of the Statistical Structure of a Random Series of Binary Symbols: H. W. HAKE AND R. HYMAN 64

American Psychological Association

Vol. 45 No. 1 January 1953

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. All of the cost of an author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas. Priority in publication is given to articles whose authors assume the full cost of publication.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 45, NO. 1 JANUARY, 1953

TRANSFER IN MOTOR LEARNING AS A FUNCTION OF DEGREE OF FIRST-TASK LEARNING AND INTER-TASK SIMILARITY [1]

CARL P. DUNCAN

Northwestern University

The purpose of this study was to investigate transfer when both the first and second tasks consisted of different patterns of S–R connections on the same apparatus. This type of transfer situation has not received much study, although it occurs frequently. Equipment requiring the operator to learn a fairly complex coordination, e.g., automobiles, airplanes, is constantly being modified. Such modifications often require the operator to learn a new S–R pattern. The present study is an attempt to provide data on transfer when one has to relearn operation of the same device because S–R relationships have been changed.

A brief description of the apparatus may help clarify the transfer situation employed. The *S* responded to colored-light stimuli by moving a lever, grasped by the right hand, into a slot. There were six slots, arranged as radii of a circle, and six lights. Learning the light associated with each slot was made difficult by requiring *S* to hold another lever, with the left hand, in a fixed position. One pattern of light-slot pairs defined Task I; Task II was provided by pairing the lights with different slots. Different degrees of inter-task similarity were obtained by varying the number of lights paired with new slots on the second task.

Few studies relate to the present approach. The most similar work in verbal learning appears to be the case in which the stimuli of two lists learned successively were identical and responses similar. This condition resulted in positive transfer (6).

Perhaps more relevant to the present study is the work of Lewis, McAllister, and Adams (3), where transfer was studied by reversing arrangement of controls on the Mashburn apparatus following practice on standard arrangement. On the reversed task *S* had to move controls in the direction opposite to that on the standard task. Some interference, in terms of errors on the first trial of reversed-task practice, was found, but over-all transfer was positive. There was some evidence that both interference and positive transfer increased with degree of first-task learning, but the relationship was not clear-cut.

The method in the present study of defining similarity between tasks in terms of number of S–R connections changed from Task I to Task II appears similar to the procedure used by Crafts

[1] This is the first of a series of papers on transfer in motor learning. The research was performed under Air Force Contract No. 33(038)-11396 between the Aero Medical Laboratory of the Wright Air Development Center, Air Research and Development Command, and Northwestern University. Thanks are due to Dr. B. J. Underwood, Assistant Contract Director, and to Dr. V. E. Montgomery, Chief Research Assistant.

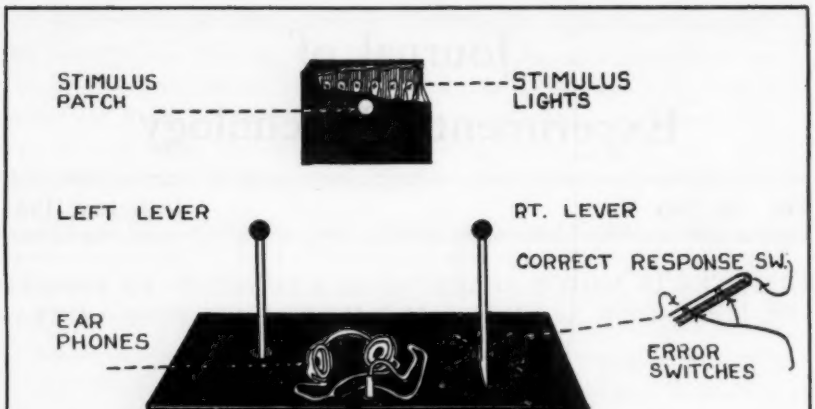

FIG. 1. The apparatus, partly schematic

(1), in an experiment on card sorting. Cards were sorted in nine boxes and Task II was provided by changing, for different groups of *Ss*, positions of 0 (control), 3, 6, or all 9 boxes. Sorting time on Trial 1 of Task II increased directly with number of boxes with changed positions. It is likely, extrapolating from the data Crafts presented, that transfer would have been positive if more trials than the two reported had been given on the final task.

The above studies suggest we may expect to find positive transfer in the present experiment and to find it increasing directly both with degree of learning of Task I and similarity between tasks.

PROCEDURE

Apparatus.—Figure 1 shows a partly schematic drawing of the apparatus. The figure shows the stimulus unit, the response unit, and an enlargement of one of the slots. The electrical connections and the recording unit are not shown.[2]

Two levers of ¼-in. steel rod projected vertically 8½ in. above the top surface of a box 37 in. long, 16 in. wide, and 8½ in. deep. The levers, topped by knobs, were 21 in. apart. The right-hand lever was centered in a circular opening, 1¼ in. in diameter, cut in the center of a steel plate, 16 in. square, that formed part of the top surface of the box. Six radial slots, 3½ in. long, .26 in. wide, and spaced 60° apart, led from the perimeter of the 1¼-in. hole in the plate. The lower end of the right-hand lever was connected to a ball-and-socket joint, permitting movement of the lever into any slot.

Each slot was equipped with three microswitches. The first was ½ in., the second 2½ in., from the entrance of the slot. The third was at the far end of the slot. With the lever in an incorrect slot a shallow error was recorded by tripping the first switch; tripping the second switch recorded a deep error. Depressing the third switch recorded a correct response. With one exception (see below) the error circuits were

[2] Wiring diagram of the apparatus is available on request.

cut out for any slot whenever it was correct for the stimulus light showing. Correct responses, shallow errors, and deep errors were recorded on separate counters.

The left-hand lever was supported below the surface of the box by a rubber pad, so it could be moved in three dimensions. Attached to the lower end of left lever was a small flexible spring projecting into a brass cup. Movements of the top of the left lever ¼ in. or more in any direction except straight up touched the spring to the cup, breaking a circuit to a .01-sec. Standard Electric clock. The left lever could not be held steady by lifting vertically because the knob, fitted on by a sleeve, would come off. Thus the left lever provided essentially a steadiness task. The clock recorded duration of time the lever was held correctly; if held incorrectly, a circuit was completed to an oscillator generating a 1000-cycle tone. The tone, delivered to *S* through earphones, informed him to correct the lever's position. The two levers were so connected that correct responses were not recorded if the left lever was out of position. At such times entry of the right lever into a correct slot yielded a deep error and the stimulus light would not change.

The box containing the levers rested on a table the height of which was adjustable by means of pulleys and counterweights.

Light sources for stimuli were 50-w. projection bulbs arranged in a semicircle so the light from each was projected onto the rear surface of a single ground-glass screen. The light from each bulb passed through a different gelatin filter to provide the six hues. Filters were made for the purpose and transmitted a fairly wide band of wave lengths. Although brightnesses and saturations undoubtedly varied, the hues could easily be differentiated.

Projection bulbs were contained in a box 18 in. long, 12 in. wide, and 9 in. high. Viewed by *S*, the stimulus box presented a black surface in the upper center of which was a circular opening, 2½ in. in diameter, covered by the ground-glass screen. The screen was 30 in. from *S*'s eyes. Stimulus hues were seen by *S* as diffuse color filling the screen.

Each of the six stimuli appeared eight times in a randomized sequence of 48 lights controlled by a rotary switch with 48 positions. The sequence repeated itself continuously. The same light never appeared twice in succession. Some *S*s went through the sequence many times but there was no evidence of memory for serial order.

A panel of six telephone jacks permitted *E* to connect any light with any slot, so a new task could be provided for *S* in a few seconds. A microswitch riding on a cam driven by a 1-rpm motor automatically provided 20-sec. trials with 10-sec. intertrial rests. The apparatus could also be set for continuous run.

Experimental design.—The two variables studied were degree of learning of Task I and similarity between tasks. Task I was learned for 10, 40, 80, or 180 trials and there were three degrees of inter-task similarity. All 12 combinations of these conditions were studied in a factorial design with 25 *S*s in each of the 12 cells. Notation that is used throughout the paper is shown in Table 1; roman numerals indicate Learning Groups (75 *S*s each), capital letters indicate Similarity Groups (100 *S*s each).

Table 1 shows number of trials given each Learning Group on Task I. All *S*s were given 60 trials on Task II. For all *S*s 2 min. elapsed between the last trial of Task I and the first trial of Task II.

Distribution of the experiment over days was as follows: Group I, all trials on both tasks on the same day; Group II, Trials 1–35 of Task I on one day, Trials 36–40 of Task I and all trials on Task II on the succeeding day; Group III, Task I Trials 1–40 on Day 1, 41–75 on Day 2, and 76–80 and Task II on Day 3; Group IV, Task I Trials 1–60 on Day 1, 61–120 on Day 2, 121–175 on Day 3, and 176–180 and Task II on Day 4. Note that in Groups II, III, and IV the last five trials of Task I were given immediately preceding the start of practice on Task II to eliminate effects of forgetting over the last 24-hr. rest, in this respect making these groups comparable to Group I where there was no 24-hr. rest.

The tasks.—Task I was defined as a particular pairing of lights with slots. Task II was a different pairing of lights and slots. Inter-task similarity was defined in terms of the *number*, out of a possible six, of light-slot connections that were *different* on Task II from what they had been in Task I. In Group A (high similarity) two lights were connected to different slots for Task II; the other four lights were paired with the same slots throughout both tasks. Group B (medium similarity) had four lights paired with different slots in Task II; Group C (low similarity) had all six lights and slots newly paired in Task II.

TABLE 1

DESIGN OF THE EXPERIMENT

Degree of Task I Learning	Inter-task Similarity		
	A(High)	B(Medium)	C(Low)
I (10 trials)	25	25	25
II (40 trials)	25	25	25
III (80 trials)	25	25	25
IV (180 trials)	25	25	25

Although it was assumed that all 720 combinations of lights with slots would be equally difficult to learn, only 25 combinations, chosen so each light was connected to each slot an approximately equal number of times, were used for Task I. Each *S* within a subgroup of 25 *S*s was assigned a different combination before practice began on Task I.

For each combination used in Task I, three other combinations were made up for use in Task II. One of these had high similarity, one medium, and one low. In choosing the Task II combinations for each Task I combination, wherever possible one of the other 24 Task I combinations was used. Thus, the Task II combination for any one *S* was often one used by a different *S* as Task I. We shall therefore assume that the tasks were not only equal in difficulty but were in a sense identical. Because of this assumption, separate control groups were not run. Instead, data from Task I were used to construct control curves for learning Task II.

Controls.—The control curve for learning Task II was constructed from data of Groups III and IV on Task I. These were the only groups given at least 60 trials on Task I, the number given on Task II.

Subjects and method.—The *S*s were 300 males from introductory psychology classes at Northwestern University. They were assigned randomly to the 12 subgroups. They were rewarded with one point of examination credit for each day served and were told the experiment was being done for the Air Force; as a result, they appeared to be highly motivated.

When *S*s first appeared for the experiment, they were asked to name the six hues as *E* presented them in the stimulus window. This was continued until *E* was satisfied *S* could easily discriminate the lights. Upon questioning, a few *S*s said they were color-blind, but if they could name the hues, they were used. One *S* was rejected for inability to discriminate the lights.

Following hue naming, *S* took his place in front of the apparatus with hands on levers. Height of the table was adjusted until knobs of the levers were approximately 1 in. below *S*'s elbow height. Since *S* stood up for all practice trials, this height seemed to be least fatiguing. The *S* was then instructed on the apparatus and what he had to do. In pilot experiments lengthy instructions had been developed, of which the gist was as follows: *S* was told the experiment concerned transfer and division of attention, the former explained with reference to different makes of cars, the latter by showing *S* he could not turn off lights unless the left lever was held correctly. The *S* put on earphones and listened to the oscillator tone. He was told to learn, by moving the right lever into slots, which light was connected to which slot, and to turn off as many lights as possible during each 20-sec. trial. He was further told that during the 10-sec. intertrial rest he would hear how many correct responses he had made in the previous trial.

Between trials, *S* remained in position at the apparatus with hands on levers and earphones on. After every block of 20 trials on the same task a 3-min. rest was given during which *S* took off earphones, sat down, and was permitted to smoke and talk.

When *S* returned on subsequent days, he was told the lights and slots were paired the same as before. Upon completion of the required number of trials on Task I, *S* was given a 2-min. rest while light-slot connections were changed for his Task II. The *S* was then instructed that 2, 4, or all 6 lights (depending upon the similarity condition) were now connected to different slots and practice was started on Task II.

Results

Comparability of groups on Task I.—Before we can present data on the comparability of groups in performance on Task I, it is necessary to show that the measures are reliable. A reliability coefficient based on all 300 *S*s was computed from the total correct responses on the first five odd and first five even trials. The coefficient was .790, corrected to .883.

Another coefficient was computed from the total of the first 20 odd and first 20 even trials, where the available *N* was 225. The uncorrected coefficient was .965. Since most of the findings are based upon correct responses as the measure, reliability coefficients were not computed for other response measures. The coefficients obtained indicate sufficient reliability to permit group comparisons.

Since we are concerned with transfer, data for acquisition of Task I are not presented in detail. To determine if random assignment of *S*s to subgroups resulted in equal performance on Task I simple analysis of variance was performed on total correct responses of the first ten trials, which

TABLE 2

MEAN SCORES ON THE LAST TWO TRIALS OF TASK I

Group	Correct Responses		Steadiness Time (Sec.)	
	Mean	σ_M	Mean	σ_M
I	4.63	.237	9.67	.479
II	10.54	.416	12.32	.472
III	13.20	.344	13.87	.369
IV	15.63	.387	14.52	.403
Grand mean	11.00	.294	12.59	.224

permits comparison of all 12 subgroups. It was first demonstrated that the 12 variances were homogeneous; Bartlett's test resulted in a chi square of 1.92 with 11 *df*. Since the variance within groups (149.62) was larger than that between groups (108.06), the subgroups appear to be adequately matched.

The three subgroups within each Learning Group were also tested for equality of performance by four separate analyses of variance. For each analysis the score was total correct responses for all trials given that particular Learning Group on Task I. The *F*'s were less than one in Groups I, II, and III. In Group IV *F* was 2.69, with 2 and 72 *df*. Since *F* at the 5% level for 2 and 80 *df* is 3.11, the subgroups within each Learning Group can be considered equal in performance on Task I.

Table 2 shows mean correct responses per trial and mean steadiness time per trial with the left lever for each Learning Group at the end of practice on Task I. Values in the table are based on the mean of the last two trials of Task I practice for each group. The values are used to

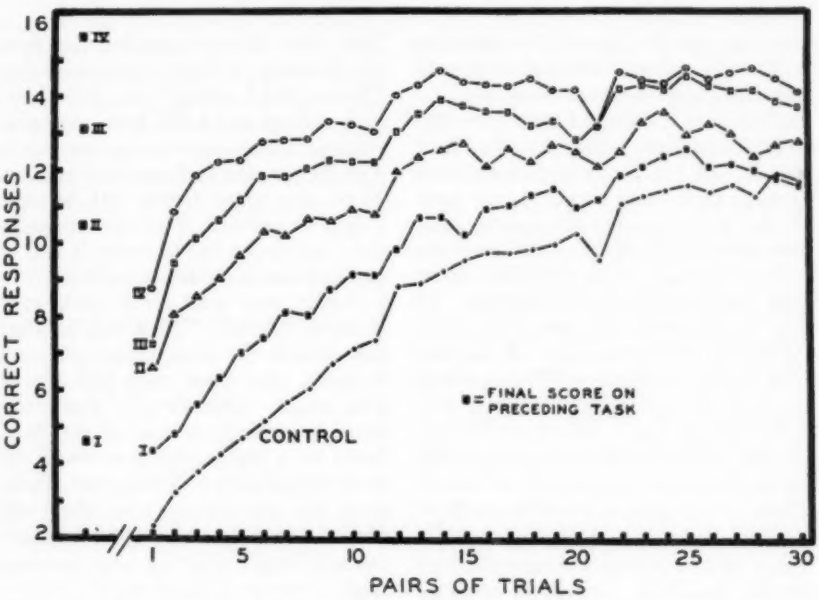

FIG. 2. Performance on Task II as a function of degree of learning of Task I with inter-task similarity held constant

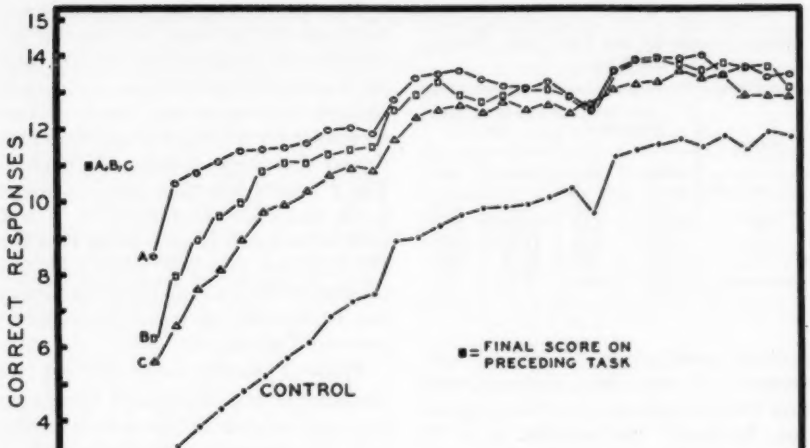

FIG. 3. Performance on Task II as a function of inter-task similarity with Task I learning held constant

measure gains or losses in transferring to Task II and will be shown on graphs depicting acquisition of Task II.

Transfer: Correct Responses.—Acquisition of Task II is shown in Fig. 2 and 3. In Fig. 2 the three Similarity Groups have been combined for each of the four Learning Groups to show the effect on Task II of different degrees of learning of Task I when inter-task similarity is held constant. In Fig. 3 Learning Groups have been combined to show effect of varying inter-task similarity. The control curve is shown in both figures, and both include heavy points, at the left of the abscissa break, representing level of mastery at the end of Task I. These heavy points were taken from Table 2.

Let us first inspect the control curve in Fig. 2 and 3. It will be recalled that this is the curve for the first 60 trials of Groups III and IV on Task I. The curve shows a sudden rise from the eleventh to twelfth abscissa points (Trials 21–22 and 23–24), and shows both a drop and a rise from the twentieth to twenty-second abscissa points. The drop is due to forgetting over the 24-hr. rest given Group III between Trials 40 and 41. The rises are probably indirectly due to rests; it will be recalled that in addition to 24-hr. rests a 3-min. rest was given after each block of 20 trials. Although we shall not present the data, there was some evidence that these rests had a complex effect. Occasionally there was some forgetting, or loss of set, indicated by a slight drop in score on the trial immediately following rest. However, on the second trial after rest there frequently was a gain (reminiscence?) larger than on any previous trial.

Referring to the curves for the experimental groups, both Fig. 2 and 3

show there is considerable initial decrement, indicated by the drop from end of Task I to beginning of Task II. A clearer picture of initial decrement may be obtained from Fig. 4, where initial loss in correct responses on Task II is plotted as a function of trials on Task I. The three curves in the figure represent similarity conditions. Values of points on the curves were computed by subtracting the mean of last two trials on Task I from the mean of first two trials on Task II; thus, a loss is shown by a minus value.

Figure 4 shows that initial loss is a negatively accelerated increasing function of Task I learning. Differences among the curves show decrement increased as inter-task similarity decreased. It is clear that initial performance on Task II is a function of both variables under consideration.

Returning to Fig. 2 and 3, we see that, regardless of initial decrement, there is no interference, at least in terms of correct responses; no curve falls below the control curve except Group I for the last four trials. The over-all transfer effect is clearly positive. No test of significance appears necessary; it can be concluded that all experimental groups showed positive transfer.

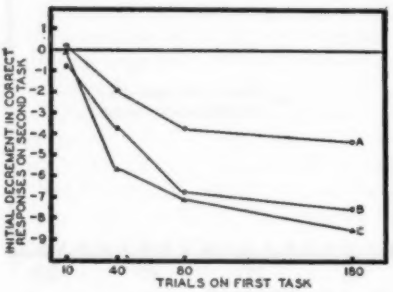

FIG. 4. Initial decrement on Task II measured as the difference between mean of last two trials on Task I and mean of first two trials on Task II

TABLE 3

ANALYSIS OF VARIANCE OF TOTAL CORRECT RESPONSES FOR ALL 60 TRIALS ON TASK II

Source	*df*	Mean Square	*F*
Learning	3	693,260.89	39.54*
Similarity	2	123,939.03	7.06**
L × S	6	7,786.30	
Within groups	288	17,532.52	

*F = 3.88 at the 1% level, 2.65 at the 5% level, for 3 and 200 *df*.
** F = 4.71 at the 1% level, 3.04 at the 5% level, for 2 and 200 *df*.

It cannot be determined from Fig. 2 and 3 whether or not the two variables interact. To assay this, double classification analysis of variance was performed on total correct responses for all 60 trials of Task II. First it was determined that the hypothesis of homogeneous variance among the 12 subgroups could not be rejected; chi square by Bartlett's test was 7.56.

The analysis of variance is summarized in Table 3. It can be seen that learning and similarity do not interact significantly. Using variance within groups as error, the F's show that both learning and similarity are highly significant.

The analysis just presented utilized all 60 trials, so it tends to mask possible changes in effectiveness of the variables over practice. Examination of Fig. 2 and 3 indicates that differential mastery of Task I may have been effective throughout all trials on Task II, whereas inter-task similarity does not appear differentially effective after Trial 30. The possibility that learning had an effect that lasted later in practice than the effect of similarity was tested by separate analyses of variance, one on total correct responses for the first 30 trials, and one on the last 30 trials, of Task II. This division of trials into first and last 30 has no theoretical basis; it was chosen arbitrarily from inspection of Fig. 2 and 3.

It was first demonstrated that variances of the 12 subgroups on the first 30 trials were sufficiently homogeneous to permit analysis; chi square was 4.59. It was assumed without testing that homogeneity of variance obtained for the last 30 trials.

We shall not present summary tables for these analyses. For both, the *df*, and thus *F*'s required for significance, are those given in Table 3. Analysis of the first 30 trials showed learning highly significant ($F = 61.11$), similarity highly significant ($F = 17.32$), and interaction not significant. For the last 30 trials learning was highly significant ($F = 18.31$), similarity not significant ($F = 1.11$), and interaction not significant. Interaction variance was less than within-groups variance in both analyses. These analyses therefore demonstrate that differential positive transfer produced by varying Task I learning extended later in practice than that resulting from varying similarity.

Transfer: Errors.—It will be recalled that deep errors were recorded both when right lever traversed completely an incorrect slot and when the slot was correct but left lever was out of position. Because the two types of error were not measured separately, the deep-error count turned out to be a rather gross and not particularly helpful measure. The shallow-error count better represented incorrect responses, since it was based only on partial entries of right lever into incorrect slots. Separate analyses were made of both error counts and of total error score obtained by combining shallow and deep errors. Results of these analyses will not be presented in detail because they yielded very little information beyond that derived from analysis of correct responses. For example, error curves for Task II also showed positive transfer; all experimental group curves were lower (showed fewer errors) than the control curve until the last ten trials (Trials 51–60), where all curves began to overlap. There were clear differences both among the three curves for Similarity Groups and the four curves for Learning Groups for about the first 30–40 trials on Task II, after which there were little or no differences in errors

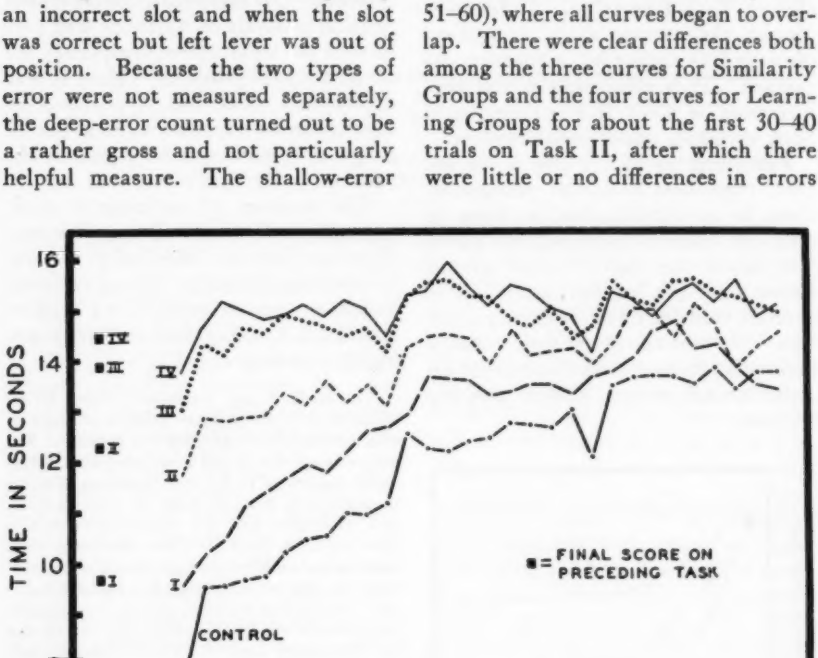

FIG. 5. The time per trial that the left lever was held in correct position during acquisition of Task II as a function of degree of learning of Task I with inter-task similarity held constant. Maximum score is 20 sec.

produced by various degrees of either variable. Finally, there was almost no evidence of interference in terms of errors. As in the case of correct responses, all groups showed considerable initial decrement (*increase* in errors) from the end of Task I to the beginning of Task II, but only Learning Group I and Similarity Group C made more errors than the control group, and this only on first two trials of Task II.

Transfer: Time scores.—The *S*'s task with the left lever remained the same throughout all phases of the experiment; viz., to hold the lever steady. The effect this might have on transfer will be discussed later. It does not seem necessary, however, to present a detailed analysis of the time score, i.e., the time per trial the left lever was held in correct position. The one graph presented, Fig. 5, shows time in seconds per trial on Task II for the Learning Groups. Similarity Groups are combined within each Learning Group. The graph also includes the control curve and the heavy points, taken from Table 2, showing the level reached at end of Task I. Figure 5 shows that there was some loss in time score, at least in Groups II, III, and IV, at start of practice on Task II; however, the decrement was regained by the fourth or fifth trial. Thus the decrement demonstrated by using correct responses (and, in fact, errors) as measures, since it lasted much later in practice than the fifth trial, was not due merely to loss of skill with the left lever. It may also be seen in Fig. 5 that Groups I and II, and perhaps Group III, continued to improve in steadiness throughout practice on Task II.

Discussion

The most striking result of the experiment was the large amount of positive transfer obtained. Every degree of first-task learning facilitated acquisition of Task II, and facilitation occurred for every variation of intertask similarity. These results corroborate those of Lewis, McAllister, and Adams (3); when stimuli remain the same from task to task and direction of movement involved in the responses is changed, positive transfer results.

It is believed that two factors are responsible for the positive transfer obtained here: response generalization and learning-how-to-learn. Concerning response generalization, Underwood (6) has argued that this factor is responsible for the positive transfer typically found in learning a second paired-associate verbal list whose stimuli are identical with, and responses similar to, those of the first list. Since he presented the theory in some detail, it is only necessary to summarize it here. The basic assumption is that reinforcement of a response during learning the first task or list simultaneously results in some strengthening of similar (generalized) responses. A gradient of response generalization is also assumed; the more similar responses are to the reinforced response the more they are strengthened. Thus, as a result of "parasitic" reinforcement of generalized responses during practice on the first list, there will be some tendency to make responses of the second list even before practice begins. The result is facilitation of second-list learning.

In Underwood's experiment response similarity was varied over several steps. In the present study whatever degree of response similarity existed, it was the same for all *S*s. (Note that we are considering response similarity, not intertask similarity as it is defined in the present experiment.) It seems reasonable to assume that the responses of the two tasks used here were highly similar since even in the group where all stimuli and responses were newly paired on Task II, only the direction of movement was

changed, not always by as much as 180°. Gagné, Baker, and Foster (2) have argued that direction is only one of several dimensions of movement similarity. Thus the present experiment may be compared with the condition of high response similarity in Underwood's study. For this condition (high response similarity) there was positive transfer to the second task in both experiments and the transfer increased sharply with degree of first-task learning. It may also be pointed out that in the Underwood experiment, although in general positive transfer increased with first-task learning at each degree of response similarity, the increase was greater and more clear-cut with high response similarity.

The other source of positive transfer is that which results from "learning-how-to-learn" (4). A practice effect in learning successive samples of the same or similar material has often been observed (for summaries, see 4 and 5). Learning-how-to-learn almost certainly occurred, since *S*s were naive to the tasks. Furthermore, *S*'s task with left lever remained the same at all times, i.e., to hold the lever correctly. Since correct responses with the right lever could not be made unless the left lever was held correctly, we may have a specific case where learning-how-to-learn was required. Differences in positive transfer resulting from different degrees of Task I learning, shown in Fig. 2, may be in part due to differences in skill with the left lever, shown in Fig. 5.

Since learning-how-to-learn was not controlled, we cannot know how much it contributed to positive transfer and how much was due to response generalization. However, learning-how-to-learn was probably not the only source of transfer, since in Underwood's experiment with verbal lists (6), where it was better controlled, considerable transfer resulted from response generalization alone.

It was noted above that our results are quite similar to those obtained by Lewis, McAllister, and Adams (3), where positive transfer was obtained when the second task was provided by reversing the direction of movement of the controls on the Mashburn apparatus. It is believed that the interpretation of positive transfer given above can, without change, also be applied to the findings of Lewis *et al.*

Variation of what may be called "formal" similarity between the tasks in the present study was accomplished by varying the number of stimuli and responses newly paired on Task II. The result was much the same as that found by Crafts (1) when he changed positions of various numbers of boxes in card sorting; as similarity decreased, positive transfer decreased. However, in the present experiment there was no evidence of interference when the measure used was correct responses; as Fig. 3 shows, even the group practicing the least similar task (Group C) was well above the control on all trials. It is true that, as reported above, there was slight interference in Group C in errors, but it did not extend beyond the third trial.

It seems important to emphasize the precise effect upon transfer obtained by varying the number of changed S–R connections; within limits of such variation obtainable here, transfer varied directly with *number* of S–R connections changed. This relationship between transfer and mere number of new S–R pairs held for all degrees of Task I learning; there was no evidence of interaction between learning and similarity. There is no a priori reason why this should be true. Without evidence to the contrary one could just as easily assume that in a complex task, where high performance depends upon integration of several functionally interrelated components, changing even a few components would disrupt performance as severely as changing several. That this did not occur in the present experiment lends some justification to analyzing even complex tasks in terms of specific stimuli and responses.

Summary

Transfer between two tasks, both available on the same apparatus, was studied as a function of two variables: degree of learning of the first task and similarity between tasks. The *S*s were required to associate each of six colored-light stimuli with movements of a lever, grasped by the right hand, into six radially arranged slots, while simultaneously holding another lever steady with the left hand. There were four degrees of first-task learning and three degrees of inter-task similarity arranged as a factorial design in which 300 *S*s were used, 25 in each of 12 cells. The degrees of learning were: 10, 40, 80, and 180 trials. Similarity between tasks was defined in terms of the number of lights newly paired with different slots on the final task; in the three similarity conditions either 2, 4, or all 6 of the lights were paired with different slots on the final task. All *S*s were given 60 trials on the final task. The results were:

1. Transfer was positive for all experimental groups whether measured in terms of correct responses or errors. Transfer increased directly both with first-task learning and inter-task similarity.

2. There was no interaction between learning and similarity.

3. Differential positive transfer resulting from varying first-task learning was obtained throughout acquisition of the final task. Differential positive transfer resulting from varying inter-task similarity occurred only during the first half of second-task practice.

4. Loss in performance from end of the first to beginning of the final task varied directly with both variables, but in all experimental groups initial performance on the final task was superior to the control.

5. There was almost no evidence of interference, measured by errors, between tasks.

6. Transfer was attributed to a combination of response generalization and learning-how-to-learn.

(Received July 17, 1952)

References

1. Crafts, L. W. Transfer as related to number of common elements. *J. gen. Psychol.*, 1935, **13**, 147–158.
2. Gagné, R. M., Baker, K. E., & Foster, H. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67–79.
3. Lewis, D., McAllister, D. E., & Adams, J. A. Facilitation and interference in performance on the modified Mashburn apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, **41**, 247–260.
4. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
5. Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.
6. Underwood, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first-list learning. *J. exp. Psychol.*, 1951, **42**, 44–53.

Journal of Experimental Psychology
Vol. 45, No. 1, 1953

THE LAW OF PRIMARY REINFORCEMENT IN CHILDREN[1]

PAUL S. SIEGEL AND JAMES G. FOSHEE

University of Alabama

In protest of strict "rat psychology," Hilgard has advanced a serious argument for more naturalistic observation and the utilization of the human subject in learning research (1). He feels that principles coming from the animal laboratory should be validated at a human level. Similar considerations have inspired several recent experiments. For example, Hubbard (2) has investigated secondary reinforcement in the human subject and Saltzman (5) has studied the effect of delayed reward. The findings of both are in accord with the results of infrahuman studies.

The present experiment was designed to examine the principle of primary reinforcement utilizing the human subject. Stated briefly, this principle maintains that the strength of a habit is positively related to the number of times it has been rewarded or accompanied by need or drive reduction. It was elected because of its central role in reinforcement-type learning theories. The experimental findings of Perin (4) and of Williams (8) are usually adduced in its support (3). Employing the rat in a bar-pressing situation, they varied the number of reinforcements and measured subsequent resistance to extinction. The relationship was found to be a positive one. Our methodology parallels theirs with one significant variation. We employed the child as a subject.

[1] This research was supported in part by a grant from the Research Committee of the University of Alabama.

METHOD

Apparatus.—The apparatus consisted of a Masonite box, painted green, of outside dimensions, height, 20 in., depth, 10 in., and width, 6½ in. In front, at the base, was a wooden tray, 3½ × 3½ in. Four inches above the tray, a ¼-in. steel rod projected ¾ in. A pressure of 30 gm. resulted in an excursion of the rod through approximately 1¼ in. This served to close a mercury switch, energizing an electromagnetic release mechanism. A single pellet of candy was released from a magazine and fed by chute to the tray. An impulse counter and a remote switch were wired in series with the release mechanism. Following the fall of each pellet, *E* opened the circuit to prevent "jiggling" and thus ensure a discrete response for each reinforcement. During the extinction series, this switch remained closed. The pellets were fashioned by splitting party mints and trimming to a cylindrical shape. Each weighed approximately .2 gm. The box was either placed upon the floor or on a nursery school table. The *S* was always free to sit or stand as he chose. The *E* was seated to one side at a distance of 3 to 6 ft. depending upon the size of the experimental room.

Subjects.—The ages of *S*s ranged from 2 yr., 11 mo. to 5 yr., 8 mo. The *S*s were divided into four groups, matched for age within a range of 3 mo., and balanced for sex. There were ten males and ten females in each of four groups (total = 80). The four groups received 2, 4, 8, or 16 reinforcements. In addition to the 80 *S*s constituting the experiment, 24 others were tried and discarded for various reasons. Nine failed to follow instructions. In 11 instances, the apparatus jammed. In 4 cases, *E* could not match for age.

Procedure.—Experimental sessions were conducted from 9:30 A.M. to 10:30 A.M. Most of the children were reported as eating breakfast at approximately 7:30 A.M., making for a food-privation interval of 2 to 3 hr. The junior author served as *E* throughout the entire experiment, making for a uniform social environment. To establish rapport, *E* devoted a minimum of 4-hr. social interaction with each nursery school or kindergarten group from which *S*s were recruited.

The experimental procedure was conducted as follows. The *E* entered the school yard and

TABLE 1

THE RELATIONSHIP BETWEEN RESISTANCE TO EXTINCTION AND NUMBER OF PRIOR REINFORCEMENTS

No. Prior Reinforcements	Number of Responses during Extinction				
	Mean	*SD*	Median	Mean Logs	*SD* Logs
2	32.9	49.8	10.0	1.12	.59
4	41.0	34.0	29.0	1.49	.84
8	51.6	33.7	49.5	1.64	.82
16	94.9	48.9	99.5	1.91	.89

inquired of a child if he cared to see a new game. If he responded negatively, no effort was made to coax him. If he indicated acceptance, he was led to a nearby experimental room and told that he could sit or stand before the apparatus. The *E* then pointed to a single pellet in the tray and asked, "What is it?" If *S* indicated recognition, he was told that he could eat it. If no recognition was apparent, *E* told the child that it was candy and that he was free to eat it. Next, *E* pointed to the lever and inquired, "What do you think would happen if you pushed this?" If *S* pressed the lever, ate the candy, and pressed again, *E* said nothing further. If *S* pressed the lever and then failed to eat the candy, *E* said, "You may have as much as you like." No further comment was offered. The *S* was given a predetermined number of reinforcements, the magazine having been loaded with 2, 4, 8, or 16 candy pellets. The extinction series was ushered in by the first failure to secure a pellet. The number of unreinforced responses made by the child in the 3 min. immediately following the last reinforced response was recorded. This is our measure of resistance to extinction.[2]

During extinction, *E* made no response to questions addressed him by the child. He simply maintained a "poker face."

RESULTS AND DISCUSSION

Table 1 summarizes our findings. The means, standard deviations, and medians for the four distributions of raw scores are found there. The relationship between the mean and the median suggested positive skewness, particularly in the instance of the two groups receiving few reinforcements (also true of the Perin-Williams rat data). Snedecor's g_1 test for skewness (6) was applied and significance at the .01 level was indicated. A logarithmic transformation of the data was made and found to be essentially normal. The means and *SD*'s of the log values are also presented in Table 1.

The data of Table 1 strongly suggest a positive relationship between resistance to extinction and number of prior reinforcements. As the number of reinforcements increases, the average number of extinction responses (both means and medians) increases. The statistical significance of this trend may be evaluated in several ways. Basically, we must assess the significance of the differences obtained among groups. We have applied two tests, one to the nonnormal raw score distributions and one to the normal or logarithmic transformations. The first, a chi-square analysis, is found in Table 2. The median of the total distribution (combining all groups, $n = 80$) was determined and a simple count of subjects yielding scores above and below this value was noted for each group. If resistance to extinction is independent of number of reinforcements, then with only chance exception, ten cases should fall above

TABLE 2

CHI-SQUARE ANALYSIS FOR DIFFERENCES AMONG GROUPS

Number of Reinforcements	Above Median of Total Group	Below Median of Total Group	*N*
2	5(10)	15(10)	20
4	7(10)	13(10)	20
8	12(10)	8(10)	20
16	16(10)	4(10)	20
	40	40	80

$\chi^2 = 14.8$
$p < .01$

[2] A simple enumeration of the number of unreinforced responses occurring during a brief interval of time has been reported by Walker (7) as a reliable index of resistance to extinction in the bar pressing of the rat.

this median and ten below within each group. The chi-square value of 14.8, corresponding to a probability value of less than .01, indicates that the hypothesis of independence is scarcely tenable. We conclude that significant differences obtain among our four groups or that resistance to extinction is genuinely dependent upon number of reinforcements. The second test, applied to the transformed distributions, consisted of calculating the regression coefficient for extinction upon reinforcement. This was found to be .0504 with a standard deviation of .00845. The t of 5.96 indicates a probability value of less than .01. Again, we conclude that the relationship is statistically significant.

In the Perin-Williams study, resistance to extinction was defined as the number of unreinforced responses yielded by the rat to a criterion of no responses within a 5-min. interval. Perin (4) has imposed an exponential function upon the median values with reasonably good agreement. This function could be fitted to our data. Unfortunately, with but four empirical points, many other general functions could be equally well fitted. Nothing seems to be gained by this procedure. We are satisfied to state that our results are in good general agreement with the Perin-Williams findings. It may be fairly said that with the child, we have confirmed the positive relationship between resistance to extinction and number of prior reinforcements. And, viewing our measure of extinction resistance as an index of habit strength, our results constitute a validation of the principle of primary reinforcement at the human level.

Summary

Employing the child as subject in a bar-pressing situation, the principle of primary reinforcement was examined. Number of reinforcements (candy) was varied for each of four groups. The number of unreinforced responses occurring within a 3-min. interval immediately following the last reinforced response was recorded. In good agreement with infrahuman studies, it was found that resistance to extinction or habit strength is positively related to number of prior reinforcements.

(Received July 7, 1952)

References

1. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
2. Hubbard, W. R. Secondary reinforcement of a simple discrimination in human beings. *J. exp. Psychol.*, 1951, **41**, 233–241.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.
5. Saltzman, I. J. Delay of reward and human verbal learning. *J. exp. Psychol.*, 1951, **41**, 437–439.
6. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Collegiate Press, 1946.
7. Walker, E. L. Variability in extinction scores in Skinner-box problems. *J. comp. & physiol. Psychol.*, 1948, **41**, 432–437.
8. Williams, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–522.

Journal of Experimental Psychology
Vol. 45, No. 1, 1953

ASSIMILATION OF INFORMATION FROM DOT AND MATRIX PATTERNS

E. T. KLEMMER AND F. C. FRICK

Human Resources Research Laboratories

Any situation involving perception may be considered in terms of the transmission of information. What is perceived may be described, not only by words and numbers related to the physical nature of the stimulus, but also in terms of an "amount of information" measure. Although this measure is usually defined in other terms, Fano (2) has shown that the information associated with any event is proportional to the negative logarithm of its probability of occurrence. Uncommon stimuli convey much information, whereas expected stimuli convey little. Since the information measure is a function of the probabilities of the stimuli and responses in the situation under study, it has units which are not specific to the particular stimulating conditions or particular responses. The measure may be applied without logical difficulty to any situation in which one is willing to identify the members of the stimulus and response classes and make some statements about their probability distributions. Whether or not the measure is useful in the analysis of human behavior remains to be proven. Early results from its application are, however, encouraging (3, 4, 5, 6, 7).

The present study attempts to do two things: first, demonstrate the application of the information measure to a simple visual perception, and, second, discover the maximum amount of information that can be assimilated using a particular visual code.

Suppose we ask: How much information do we take in, or how much do we learn, about spatial location in a single glance? The way we have gone about answering the question is to give a group of *S*s a very quick look at one or more dots on a plane and ask them to reproduce the location of the dots. We then applied the information measure to the presented stimulus and *S*'s responses and came out with at least a partial answer to our question.

Method

Apparatus.—The visual stimuli were photographed on 2 × 2-in. slides and projected on a screen from a 300-w. slide projector fitted with a .03-sec. shutter. The slides consisted of one or more white dots inside the white outline of a square on a black background. For some of the tests the square contained internal grid lines which divided the square into a square matrix. The dot(s) were always located in the center of the cells of the matrix. When the internal grid lines of the matrix were not included on the slide, the dots were positioned in the centers of imaginary cells of a matrix whose order was equal to the order of the matrix on *S*'s answer sheet. The *S*'s responses in each case were made on answer sheets printed with appropriate matrix patterns.

Subjects.—Undergraduate students at the University of Maryland served as paid *S*s. The same *S*s were used for the first four tests. A different group was used for the last eight tests.

Procedure.—Fifteen to 25 *S*s were run simultaneously in a semidarkened room. The *S*s faced a 40 × 40-in. projection screen which was slightly larger than the projected outline of the square. A "test" is defined as a series of slides of the same matrix order and a defined number of dots. Before each test, *S*s were informed as to the number of dots which might appear and whether or not the internal grid lines would appear on the screen. They were told that the dot(s) could occur in any position and that all positions were equally likely. During the test each possible position of the dot(s) was presented once. The order of presentation was random. The *S*s were not told, and showed no sign of learning, that no stimulus occurred twice in the same test. They

TABLE 1

CONDITIONS OF TESTS AND NUMBER OF *Ss* EMPLOYED

Test	Matrix Order	Number of Dots	Number of Alternatives	Grid Lines on Slides	*Ss*
1	3×3	1	9	Yes	26
2	4×4	1	16	Yes	28
3	6×6	1	36	Yes	28
4	8×8	1	64	Yes	22
5	10×10	1	100	Yes	42
6	10×10	1	100	No	79
7	10×10	1	100	No*	80
8	20×20	1	400	No	80
9	3×3	3	84	No	87
10	3×3	4	126	Yes	80
11	3×3	4	126	No	80
12	3×3	1–4	255	No	79

* Also no grid lines on answer sheets.

were instructed to mark the position of the dot(s) on the answer-sheet grids in the same relative position as the dots on the projected square. The exposure time was held constant at .03 sec. throughout the experiment. At least two practice slides were given before each test and a verbal ready signal was given before each flash.

The basic experimental design was as follows: In the first experiment the information content of the single-dot slides was increased by increasing the order of the matrix used to position the dot. Thus, for one dot on the 3×3 matrix there were nine possible positions. Each slide constituted a single stimulus from, in this case, a set of nine alternative stimuli. For the 4×4 matrix there were 16 possible stimuli, and so on.

In the second experiment the matrix order was held constant at 3×3 and the number of dots varied from one through four. This was done both by combining all of the one- through four-dot slides into a single test with random ordering of number as well as position of the dots and also by showing the four-dot slides and the three-dot slides as separate tests. (The one- and two-dot slides on a 3×3 matrix are received perfectly whether shown alone or in combination with the three- and four-dot slides.)

Two pairs of tests were run in which the members of each pair differed only in that one had the internal grid lines included on the projected slide and one did not. For one test there were no internal grid lines on either the slides or the answer sheets. The dot was positioned on the slide in the center of a cell of an imaginary 10×10 matrix. The *Ss* made small marks on answer-sheet squares trying to locate the dot as nearly as possible in the same relative position as the dot on the slide. The answer sheets were scored by superimposing a 10×10 matrix over the squares and tabulating the cells into which *Ss*' marks fell.

Table 1 presents the several orders of matrices, the number of dots, the grid conditions, and the number of *Ss* for each test.

Scoring.—All of the tests were scored in terms of the "Amount of information" transmitted. The information content of each stimulus is simply the negative logarithm of its probability of occurrence. If logarithms to the base two are used, we define a unit which has been termed the bit (2, 8).

In order to compute the information successfully transmitted, distributions were obtained of the responses of all *Ss* for each stimulus separately. The uncertainty associated with these distributions (sometimes called response equivocation) was then estimated from the relative frequencies of the observed responses. The average response equivocation was subtracted from the information content of the stimulus to get the transmitted information. This method assumes that the total informational content of the response is equal to that of the stimulus. This assumption was checked in several of the tests and found to be true within the limits of accuracy of these experiments. For a fuller discussion of the methods for computing transmitted information the reader should see Garner and Hake (4).

RESULTS

Before considering the effect of changing the number of cells or number of dots (increasing the informational input) let us look at the effect upon the information transmitted of including or omitting the internal grid lines. The results from com-

TABLE 2

EFFECT OF INTERNAL GRID LINES UPON THE TRANSMISSION OF INFORMATION ABOUT THE POSITION OF DOTS IN A SQUARE

Test	Matrix Order	Number of Dots	Grid Lines		H_t*
			Slides	Answer Sheets	
5	10×10	1	Yes	Yes	4.4
6	10×10	1	No	Yes	4.4
7	10×10	1	No	No	4.4
10	3×3	4	Yes	Yes	6.6
11	3×3	4	No	Yes	6.6

* Information transmitted in bits per stimulus.

parable tests, with and without such grid lines, are presented in Table 2.

From these results, which are representative of the entire range of tests in this study, it was concluded that inclusion or omission of grid lines on the slides had no effect on the amount of information transmitted. Thus, tests are compared along other dimensions without reference to grid lines on the slide.

The effect upon information transmitted of varying the number of cells in the matrix is shown in Table 3 for the single-dot tests. Increasing the number of cells in the matrix increases the number of alternative stimuli and therefore increases the amount of information presented.

It appears from Table 3 that the information transmitted increases with the information presented, but only to about 4.4 bits, where it becomes constant. This suggests that, in a single glance, we take in approximately 4.4 bits of information about the location of a point in a plane.

The results of the multiple-dot tests with the 3 × 3 matrix are given in Table 4. Only the three- and four-dot tests were run as separate tests,

TABLE 3

TRANSMISSION OF INFORMATION ABOUT THE POSITION OF A SINGLE DOT IN A MATRIX VISUALLY EXPOSED FOR .03 SEC.

Test	Matrix Order	Information in Bits per Stimulus	
		Presented	Transmitted
1	3 × 3	3.2	3.2
2	4 × 4	4.0	3.9
3	6 × 6	5.2	4.4
4	8 × 8	6.0	4.5
5,6,7	10 × 10	6.6	4.4
8	20 × 20	8.6	5.0*

* Value is inflated by small sample bias. Number of *Ss* (80) was small compared to number of alternatives (400).

TABLE 4

TRANSMISSION OF INFORMATION ABOUT POSITION OF ONE THROUGH FOUR DOTS IN A 3 x 3 MATRIX VISUALLY EXPOSED FOR .03 SEC.

Test	Number of Dots	Number of Alternatives	Information in Bits per Stimulus	
			Presented	Transmitted
1	1	9	3.2	3.2 3.2*
—	2	36	5.2	5.1*
9	3	84	6.4	6.1 6.2*
10,11	4	126	7.0	6.6 6.7*
—	1–2	45	5.5	5.4*
—	1–3	129	7.0	6.9*
12	1–4	255	8.0	7.8

* Computed from Test 12.

since the one- and two-dot messages are received perfectly, or almost perfectly, by all *Ss*. Note that the number of alternative stimuli and, therefore, the average information presented goes up rapidly as the number of dots is increased. Combining several numbers of dots into a single test added alternative stimuli and further increased the information presented.

Only the values in the first information-transmitted column were obtained from separate tests. However, as the results of the three- and four-dot tests clearly indicate, there was no loss in accuracy of response when these stimuli are included in the one-to-four-dot test. This is an interesting finding. Adding the three-, two-, and one-dot stimuli to the four-dot test did not affect the responses to the four-dot messages but it did double the number of alternatives and thus add one bit to the information-carrying capacity of the four-dot messages.

It is possible to consider the one-, two-, and three-dot messages as special cases of the four-dot series. In general, n-dot messages may be assumed to include all messages from one to n. Following this reasoning the one-to-four-dots test is scored in parts as a

two-dot test which includes the one-dot slides and a three-dot test which includes the one- and two-dot slides. The result of this scoring method is shown in Table 4. Since the response to any of these slides does not seem to be a function of the company it keeps, it may be concluded that, in this case, reducing the number of alternatives by selecting only some of them for use is an inefficient means of encoding. In other words, using only the four-dot slides is inefficient when adding the one-, two-, and three-dot slides increases the information-carrying capacity of the system without increasing the error.

Note that there was very little loss of information in any of the multiple-dot tests. It is probable, therefore, that the value for information transmitted for each number of dots is below *S*'s maximum capacity. Nevertheless, it is clear that increasing the number of dots from one through four does increase the information-carrying capacity of the message from the 4.4 bit maximum for one dot to at least 6.6 bits for the four-dot case or 7.8 bits for one-to-four-dots. This is the second finding of this experiment. Increasing the number of dots from one to four increases the maximum amount of information transmitted.

Discussion

In order to get an idea of what the 4.4- to 7.8-bit range found in this experiment means, let us compare it to the information assimilated from other types of messages. An *S* can reproduce without error three to four randomly chosen letters when they are seen for .01 sec. (1). This represents an information transfer of 14.1 to 18.8 bits. Under the same conditions *S* can read about two unconnected words (1). On the assumption that the words are drawn at random from a vocabulary of 10,000 this represents 26.6 bits. It is possible that pictures would show an even higher value of information assimilated but no satisfactory measurements have yet been made.

At the other extreme, if *S* is asked to locate a single point on a straight line, the information successfully transmitted is in the neighborhood of 3 bits (5).

It is clear that the maximum amount of information that can be assimilated from a brief visual exposure is a function of the type of encoding used. The question immediately arises as to whether or not there is a common metric which may be applied to the different message classes and which will correlate with the maximum information-carrying capacity of that class.

Of course, we are not in a position on the basis of these scanty data to draw general conclusions applicable to other forms of encoding. However, there is a suggestion here. It takes two coordinates to specify the position of a single dot in a plane. It takes four coordinates to specify two dots, six for three dots, etc. These coordinates may be considered as dimensions of a message space in which each stimulus is a single point. It requires only one coordinate, or dimension, to specify the position of a point on a line. This latter type of stimulus has a maximum information-carrying capacity of about 3 bits. We have seen that the two coordinates of the single dot on a plane can transmit about 4.4 bits and the eight coordinates of one-to-four-dots at least 7.8 bits. Thus it appears that increasing the dimensions of the message space increases the information-carrying capacity of the stimulus, in this range at least. We hasten to add that there is no certainty that the message dimensions are inter-

changeable, i.e., that only the *number* of dimensions is important.

Summary

An experiment was designed to demonstrate the application of information measure to a simple visual perception and to find the maximum information about a two-dimensional spatial location which can be assimilated from a brief visual exposure. The stimuli consisted of one or more dots within the outline of a square. This pattern was flashed on a projection screen for .03 sec. The *S*s reported the positions of the dots by making marks in corresponding positions on answer sheets printed with the outline of a square and containing internal grid lines. The dots were positioned on the projected slide in positions corresponding to centers of cells in the *S*'s answer matrix. Using a single dot, the order of the matrix on the *S*'s answer sheet was increased from 3 × 3 to 20 × 20 in separate tests without changing the outside dimensions of the square. Using a 3 × 3 matrix the number of dots on each slide was increased from one through four in separate tests. The data were scored in terms of the amount of information successfully transmitted by *S* from the screen to his answer sheet.

1. The *S*s transmitted a maximum of 4.4 bits of information about the position of a single dot in a square from a .03-sec. exposure.

2. The information transmitted increased with the number of dots used to at least 6.6 bits per exposure for four dots.

3. Combining the one-, two-, three-, and four-dot slides into a single series did not result in less accuracy than when each number of dots was shown separately. Because of the increased number of alternatives, however, the information transmitted increased to 7.8 bits per exposure.

4. Including internal grid lines on the projected square had no effect on the information transmitted.

5. Removing the internal grid lines from *S*'s answer sheet had no effect upon the information transmitted by a single dot.

(Received June 23, 1952)

References

1. Cattell, J. McK. Ueber die Trägheit der Netzhaut und des Sehcentrums. *Philos. Stud.*, 1886, **3**, 94–127.
2. Fano, R. M. The transmission of information. Research Laboratory of Electronics, Mass. Inst. of Tech., Cambridge, Mass., Tech. Report No. 65, 1949.
3. Frick, F. C., & Miller, G. A. A statistical description of operant conditioning. *Amer. J. Psychol.*, 1951, **64**, 20–36.
4. Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
5. Hake, W. H., & Garner, W. R. The effect of presenting various numbers of discrete steps on scale reading accuracy. *J. exp. Psychol.*, 1951, **42**, 358–366.
6. Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the test materials. *J. exp. Psychol.*, 1951, **41**, 329–335.
7. Miller, G. A., & Selfridge, J. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.
8. Shannon, C. A mathematical theory of communication. *Bell Syst. tech. J.*, 1948, **27**, 379–423; 623–656.

Journal of Experimental Psychology
Vol. 45, No. 1, 1953

DRIVE CONDITIONING AS A FACTOR IN LATENT LEARNING[1]

PAUL E. MEEHL AND KENNETH MacCORQUODALE

University of Minnesota

In 1929 Blodgett (**1**) reported a now-classic experiment in which rats which had been running a maze without food reward exhibited a sudden drop in errors on the day following a single rewarded run. We do not propose here to discuss the experimental reality of this "Blodgett effect," which has not as yet been shown to be consistently reproducible (**2, 9, 11, 16**), nor the theoretical and statistical controversies involved in its interpretation, which have been treated at length by Kendler (**6**), Maltzman (**8**) and Thistlethwaite (**12, 13**) in recent papers. Assuming for present purposes that the effect *does* occur, we wish to examine its theoretical implications from another angle. Critics of S–R-reinforcement theory from Tolman on (**14,** p. 343; **15,** p. 194) have taken it for granted that the *feeding in the goal box* is what brings about the subsequent sudden appearance of an error decrement. But actually the experiment does not establish this, since no control feeding outside the goal box or temporally separated from a maze run was employed. From the human observer's viewpoint (especially if he has cognitive-map leanings) it is quite natural to assume that if a feeding *in the goal* and *following a run* yields the Blodgett effect, this particular spatio-temporal locus is what does the business. But such an assumption clearly requires scrutiny.

This is the more so because there is an intriguing S–R-reinforcement possibility which presents itself as soon as the above issue is once raised. It is admitted on all sides that *some* decline in errors is shown by the nonrewarded rats during the "latent phase." We need not consider here what rewards may be operative to produce this already *manifest* learning (but cf. **7**). The sticker for a reinforcement theorist, at least as we perceive it, hinges on the quantitative question, "Why such a drop from one reinforcement at this level of the learning curve?" Now one of the ways to get sudden shifts in *performance* without equally sudden changes in *habit* (or cognitive!) variables, according to S–R-reinforcement theory, is to alter *motivation*. Discussions of the Blodgett effect tend to assume that the place to look for the explanation is in the operation of reinforcement and resulting habit strength. But if in this design there is inherent some mechanism that provides for a shift in the rat's degree of hunger, might this not be a better place to look? If ${}_{s}E_{R}$ is a multiplicative function of ${}_{s}H_{R}$ and $f(D)$, increasing either factor will increase ${}_{s}E_{R}$.

Human introspection suggests that the "subjective state" of hunger, for a given interval of deprivation, can be manipulated by exteroceptive cues. Presumably Hull's Corollary i of Postulate III in the 1949 set, "Secondary Motivation" (**4,** p. 175), is intended to cover this kind of *environmental control over drive level itself*. There is independent evidence (**3,** p. 33) that some of the presumed "components of being hungry," such as stomach contractions, salivation, damped ingestive responses, and even

[1] This study was made possible through a research grant by the Graduate School of the University of Minnesota.

blood-sugar level can be brought under some degree of exteroceptive control. Thinking along these lines, we were struck by the following hypothesis: During the latent phase, differential habit strengths are accumulating in favor of the "correct" choices (as the data show). At the end of the first rewarded run, the rat feeds in the goal box and by the above corollary acquires a conditioned hunger drive (or, more precisely, his hunger drive, which during the latent runs has been based upon mere deprivation, is now strengthened by the new exteroceptive conditioning). On the following day's run, this conditioning generalizes to the similar (and partially overlapping) stimulus complex of being "in the maze," so that the rat, although under the same deprivation as on previous days, is now "hungrier" than he was on these days. Hence, we have an augmented drive factor which, according to Hullian theory, will multiply by the (also augmented!) habit strengths already favoring the correct path and hence produce the Blodgett drop.

Admittedly this is a touchy hypothesis to test, and we must emphasize that the design which follows is not presented as an *experimentum crucis* to clinch any such argument. As usual in such "derivations," the hypothesis does not actually quantify the parameters but it does require that certain rather complex relations obtain among their magnitudes. We have tried to test just *one* implication of such a hypothesis, and this test can only be added to the other numerous lines of evidence and theory needed to explicate Blodgett's experiment. However, the prediction confirmed is somewhat paradoxical and its rather bizarre character perhaps makes its success more probative than would ordinarily be the case.

For if the above hypothesis were correct, one should be able to produce the drop in errors by a feeding *not* in the goal box and *not* at the conclusion of a run. In fact, a feeding experience administered in a sufficiently "maze-like situation" should bring about the desired drive conditioning. (We are, of course, *not* talking here of "pre-feeding," i.e., just before a day's run, as a facilitator.) Accordingly, we determined to substitute a feeding experience in a maze-like but extra-maze box for the single rewarded run introduced by Blodgett, meanwhile avoiding the possibility of a directional or locational learning by actively opposing any such factor through the arrangement of apparatus.

Method

Subjects.—Thirty male albino rats, approximately 90 days old and experimentally naive, were used. These were subsequently subdivided into two matched groups of 15 rats each, as described below.

Apparatus.—We used a replica of the six-unit multiple-T Blodgett maze; unlike his, ours had a permanent floor. The maze was placed on the floor, diagonally across the experimental room, which was square, windowless, otherwise unfurnished, and quiet. The illumination came from a 7½-w. overhead light, placed symmetrically with respect to the maze and providing a uniform brightness of less than ½ ft.-candles at maze level.

The extra-maze box was a 2-ft. long section of alley identical in composition, color (black), and in height and width dimensions with the alleys of the maze. This was placed on a stand in the experimental room, so that its floor was 4 ft. above the level of the maze floor, about one yard in front of the starting box of the maze in the opposite direction from the goal box. The long axis of this box was at right angles to the major (start-to-goal) axis of the maze. This is the box in which half the rats were to be fed after a series of nonrewarded maze runs.

Procedure.—Test breaking: Since the maze trials in this experiment were to be run under conditions of food deprivation but without primary food reinforcement in the maze, we took special precautions to assure that emotionality in the experimental situation would be minimized, and that a feeding rhythm would be established, during the test-breaking phase. This lasted seven days. Throughout this period each rat

was handled and carried about for 1½ min. daily. Each rat was given about 10 gm. of wet mash daily at the hour at which he would be fed during the experimental sessions proper. Feeding was done in special feeding cages of different construction and dimensions from the home cages. On the fourth, fifth, and sixth test-breaking days, the rats were placed, in squads of five, in a tunnel-type straightaway for 20 min. The long axis of the straightaway was at right angles to the long (start-to-goal) axis of the maze. The point of entry into the straightaway was randomly varied; the point of removal was always from the middle two-thirds of the alley. The straightaway had no goal boxes and never contained food. From the straightaway the animals were removed to individual "detention" cages, differing markedly from both home and feeding cages, for 20 min., thence to the feeding cages where food was withheld for 40 more minutes, e.g., for one hour after removal from the straightaway. The purpose of the detention cages was to reduce, or at least temporally delay, the secondary reinforcing effects of the feeding cages. On the seventh test-breaking night the straightaway was not used, but instead the rats were placed singly in the extra-maze box (as described above) for 3 min., then removed to the detention and feeding cages as before.

Maze running: All rats were treated identically during the first five days. One maze trial was given daily, under 22½-hr. food deprivation. Immediately preceding his day's run, each rat was placed for 30 sec. in the extra-maze box, from which he was removed directly to the starting box of the maze. There was no food in the goal box of the maze at any time. Retracing was prevented by swinging doors. The rats were detained in the goal box 1 min. following the run. Then, as during the test-breaking period, they were removed to the detention cages for 20 min., thence to the feeding cages, where food was withheld for at least another 40 min. Records were kept of running times and of calibrated errors; an error score of 1 was recorded if the rat placed his head and forepaws but not his hind feet into the cul; a 2 was recorded if his body, including the hind feet, entered the cul but his nose did not pass the point marked 3½ in. from the end of the cul; a 3 was recorded if his nose passed this mark. According to this system, only cul penetrations to the depth scored 2 or 3 would have been counted by Blodgett as errors; henceforth such errors will be called Blodgett errors. After the fifth night's runs, two pairwise matched groups of 15 rats each were formed. The matching was on the basis of several error criteria: total number of Blodgett errors over all five days; summed weighted errors over all five days; total Blodgett errors on Day 5; summed weighted errors on Day 5. The closeness of the matching may be inferred from the failure of the curves in Fig. 1 and 2 to diverge markedly over the first five trials, and from their convergence on Trial 5.

On the sixth night, neither group was run in the maze. The rats were handled in the usual order, but those which were assigned to the "extra-maze fed" group were placed in the extra-maze box where, for the first time, they found a cup of wet mash, which they were permitted to eat for 2 min. The food cup was located about two-thirds of the alley length from the end at which the rats were inserted. The "not-fed" animals were also placed in the extra-maze box for 2 min., with no food present. All rats were removed from the extra-maze box to the detention and feeding cages as before. Each extra-maze-fed rat was ultimately given the same food cup he had eaten from in the extra-maze box, to ensure that this group did not eat more than the nonfed group. If, then, the drive-conditioning mechanism does occur as a result of eating in a similar situation, its effects should be observable in the greater error reduction of the extra-maze fed group, as compared with the nonfed group, on the subsequent run.

The sixth, or test, trial was run the following night. As before, rats were placed for 30 sec. in the extra-maze box (no food or food cup being present), and then performed one unreinforced maze run. Errors and time were recorded as before.

Results

Since there was a definite antecedent expectation that the box-fed rats would do better on the test run than

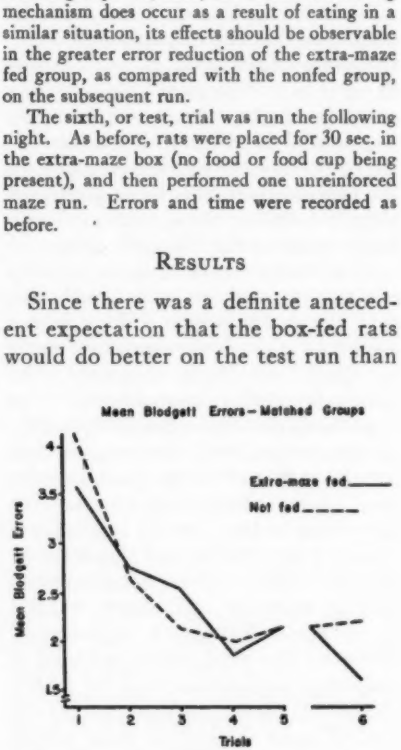

Fig. 1. Performance of two groups in terms of "Blodgett errors"

the nonfed, it seems appropriate to employ one-tailed significance tests throughout (5). Figure 1 shows the performance of the two groups in terms of "Blodgett errors" as defined above. The t test for matched groups on the test run is 2.01 ($.02 < p < .05$). For eight of the matched pairs the fed were superior to the nonfed, while for two pairs the reverse obtained. The other five pairs were equal. It should be noted that 28 of the 30 Blodgett scores lie in the narrow range of scores from 1 to 3, so this test is perhaps not very sensitive.

A more sensitive measure of actual performance is shown in Fig. 2, plotting the daily "weighted errors" as defined above. Here also the t test is 2.01 ($.02 < p < .05$). Inspection of the distributions shows that 73% of the nonfed exceed the median error score of the fed, and 73% of the latter fall below the former's median. One-third of the nonfed made more errors than any of the fed. Of the 15 matched pairs, 10 show a difference score favoring the fed rat, 1 pair is equal, and 4 show the nonfed superior. Examining these discrepancies by means of Wilcoxon's nonparametric matched pairs signed rank test (**10,** p. 126) we find a T of 27.5 as the sum

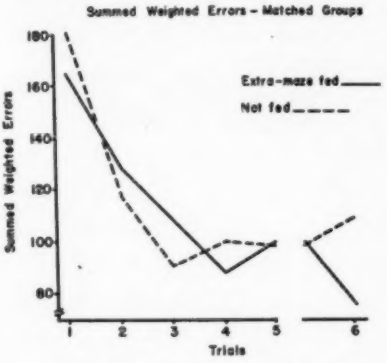

FIG. 2. "Weighted errors" of two groups

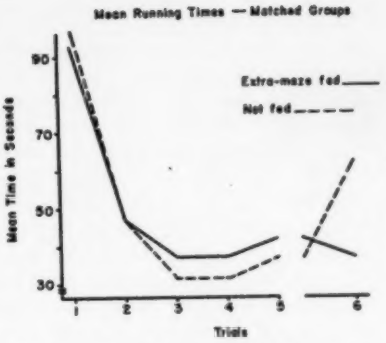

FIG. 3. Effect of extra-maze feeding upon running time

of ranks with less frequent sign, contrasted with a chance expectancy of 60 ($p < .05$; **18,** p. 120). The median test, which does not take advantage of the experimental matching, yields a p between .02 and .05 (**10,** p. 125).

An analysis of the individual rats' *gains* in weighted errors between the last prefeeding run and the test run was almost exactly at $p = .05$ ($t = 1.73$, 14 *df*) for Blodgett errors, and between the 3 and 4% levels of confidence for weighted errors ($t = 1.996$). Wilcoxon's test again yields a sum of 27.5 ($p < .05$).

The effect of the extra-maze feeding upon running times was somewhat more striking. Seventy-three per cent of the nonfed exceed the median running time of the fed, and all of the fed fall below the nonfed median. The mean running time for the two groups differed by 25.2 sec. (36.1 for the fed, 61.3 for the nonfed). A t test (matched formula) yields $t = 3.62$ ($p < .01$). Wilcoxon's test shows a $T = 10$ instead of the theoretical 60, which is at the 1% level of confidence. It seems appropriate to conclude that the extra-maze feeding exerted a significant error-reducing and speeding-up effect upon maze performance during the next day's run.

Summary and Conclusions

Hungry rats were run in the six-unit Blodgett maze one trial per day for five days, no food reward being given. During this period they showed a reduction in errors which had leveled off by the fifth day. Two groups were then formed by matching individual rats by pairs in terms of several measures of their performances to that time. On the following day none of the subjects ran the maze, but were placed in a familiar mazelike box located several feet above the maze and behind the maze end opposite the goal box. One group fed in this box on that day, the other did not. On the following day's run the fed rats were significantly superior to the nonfed in both time and errors. The results were those predicted from a hypothesis that the Blodgett effect is at least partly due to conditioning of hunger drive to the exteroceptive stimuli of the maze situation.

(Received July 10, 1952)

References

1. Blodgett, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univer. Calif. Pub. Psychol.*, 1929, **4**, 113–134.
2. Blodgett, H. C. Reynolds' repetition of Blodgett's experiment on latent learning. *J. exp. Psychol.*, 1946, **36**, 184–186.
3. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
4. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173–180.
5. Jones, L. V. Tests of hypotheses: One-sided vs. two-sided alternatives. *Psychol. Bull.*, 1952, **49**, 43–46.
6. Kendler, H. H. Some comments on Thistlethwaite's perception of latent learning. *Psychol. Bull.*, 1952, **49**, 47–51.
7. MacCorquodale, K., & Meehl, P. E. On the elimination of cul entries without obvious reinforcement. *J. comp. physiol. Psychol.*, 1951, **44**, 367–371.
8. Maltzman, I. The Blodgett and Haney types of latent learning experiment: reply to Thistlethwaite. *Psychol. Bull.*, 1952, **49**, 52–60.
9. Meehl, P. E., & MacCorquodale, K. A failure to find the Blodgett effect, and some secondary observations on drive conditioning. *J. comp. physiol. Psychol.*, 1951, **44**, 178–183.
10. Moses, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.
11. Reynolds, B. A repetition of the Blodgett experiment on latent learning. *J. exp. Psychol.*, 1945, **35**, 504–516.
12. Thistlethwaite, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, **48**, 97–129.
13. Thistlethwaite, D. Reply to Kendler and Maltzman. *Psychol. Bull.*, 1952, **49**, 61–71.
14. Tolman, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.
15. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55**, 189–208.
16. Tolman, E. C., & Honzik, H. C. Introduction and removal of reward, and maze performance of rats. *Univer. Calif. Publ. Psychol.*, 1930, **4**, 257–275.
17. Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bull.*, 1945, **1**, 80–82.
18. Wilcoxon, F. Probability tables for individual comparison by ranking methods. *Biometrics*, 1947, **3**, 119–122.

Journal of Experimental Psychology
Vol. 45, No. 1, 1953

NUMBER OF COMMON ELEMENTS AND CONSISTENCY OF REINFORCEMENT IN A DISCRIMINATION LEARNING TASK [1]

ROBERT STANTON FRENCH

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

In this study discrimination learning is investigated as a joint function of (*a*) the degree to which stimuli within a group are similar, and (*b*) the consistency of the reinforcement of a given response to these similar stimuli. These variables were chosen because it is believed that in their interaction they constitute primary determinants of learning. Results reported from a rapidly expanding area of research have given abundant confirmation to the proposition that with an increase in number of common elements, or an increase in the proximity of stimulus values along either a physical or response-defined dimension, there is almost invariably an increase in frequency of occurrence of a common response.

Equally well established is the principle that a more difficult task confronts the individual when a common response is not reinforced or otherwise encouraged, but instead a number of *different* responses are required. Simply stated, the learner must ignore the obvious similarities between the stimuli while discovering in what respects the stimuli may be differentiated. Clearly, as the number of different responses required increases, or as the proportion of similar stimuli to be associated with each of two classes of response approaches one-half, the difficulty of the task should increase proportionately. Thus, task difficulty should always increase when consistency of reinforcement, defined in either of the above ways, increases (assuming a degree of similarity among the stimuli great enough to produce generalization).

One cannot, however, describe the effect of varying stimulus similarity with such a simple statement, since there is every reason to believe that the effect of varying similarity will be markedly different depending upon the consistency of reinforcement. Considering the extremes, if a common response is reinforced for all of a group of similar stimuli, increasing the degree of this similarity clearly simplifies the task. On the other hand, if two classes of responses are involved, and if half of the similar stimuli are to be associated with each response, increasing the degree of stimulus similarity should have the opposite effect; likewise, intermediate effects should be produced by intermediate levels of the two variables. Note that a set of converging functions is being described with stimulus similarity on the abscissa, and consistency of reinforcement as a parameter. According to the hypothesis, as the degree of stimulus similarity decreases, the range of effect produced by changes in consistency of reinforcement decreases likewise, until finally with little or no similarity among the members of the "class" of stimuli, variations in consistency of reinforcement produce no measurable changes in performance.

Perhaps the largest share of research

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

concerned with stimulus similarity in its relation to learning has been in the area of transfer of training. The resemblance of the present hypothesized interaction to the change in transfer effect from positive to negative under varied stimulus and response similarity conditions should be evident. We shall consider in detail in a later section the theoretical implications of this apparent equivalence.

In the area of discrimination learning and concept formation there has been perhaps less systematic manipulation of stimulus similarity and consistency of reinforcement. A typical problem of discrimination learning involves the discrimination between two classes of stimuli with only one of the classes "correct" or associated with reward. If a common response (perhaps choosing a form or object either literally or symbolically) is reinforced for all of the members of one of these classes of stimuli (and by implication a common response of avoidance is reinforced for the other class), then the consistency of reinforcement is clearly maximum. Perhaps the simplest and most effective way of reducing this consistency by degrees to a minimum is to interchange the reinforcement or reward value of an increasing proportion of the stimuli from the two classes until half of each class is correct and half incorrect. For present purposes this latter condition defines minimum consistency of reinforcement.

Certain aspects of the present experimental hypothesis were tested by French (**2**). In his study the stimuli consisted of various combinations and permutations of four-letter aggregates of consonants. Degree of stimulus similarity was varied by manipulating the number of common letters within each of two classes of stimulus items. In addition to five degrees of stimulus similarity, two experimental conditions were contrasted which in their assignment of correct and incorrect stimulus items duplicated exactly the conditions described as maximum and minimum consistency of reinforcement for the present study. The possibility of intermediate degrees of consistency was not appreciated at the time. In brief, the results of this study clearly substantiated the hypothesized interaction.

There seems little reason to doubt that these two independent variables play a major role as determinants of learning, at least in the kind of discrimination task described. The present study is a repetition of the first in some respects, and as such can offer important corroboration. In addition, however, the present study permitted measurement of response latencies as well as errors, investigated intermediate values of consistency of reinforcement, and utilized abstract forms rather than letters. The use of abstract forms would seem to offer greater possibilities of generalizing the results to a variety of configurations than is possible with verbal materials.

By way of summary, the experimental hypothesis to be investigated may be restated briefly: It is proposed that discrimination learning performance is a joint function of (*a*) the number of common stimulus elements within a group of forms, and (*b*) the proportion of these similar forms for which the response of "choice" is reinforced or indicated as "correct." According to the hypothesis, with high consistency of reinforcement, increasing the number of common elements decreases the difficulty of the discrimination task progressively. However, as the consistency of reinforcement is reduced, the function describing decreasing difficulty changes slope progressively and is finally

replaced by a function describing increasing difficulty. Thus, in effect, an interaction between the two independent variables is hypothesized.

Method

Apparatus.—In order to test the hypothesis under consideration a technique was needed for presenting pairs of forms repeatedly in a varied order to a group of *S*s, with measurement of both the kind of response and its latency. These needs were met through the use of a projection apparatus with electric controls so arranged that four clocks started simultaneously at the moment of presentation of each stimulus pair upon the screen. Each of four *S*s, seated in a row before the screen, had a small covered box on the arm of his chair. This box contained two microswitch buttons. The *S* was instructed to indicate his choice by depressing the button on the side corresponding to the position of one of the stimuli appearing side by side on the screen. Depression of either key stopped the corresponding clock and illuminated one of two colored lights on *E*'s control panel. Six seconds after the pair of stimuli appeared on the screen (an interval sufficiently long that all *S*s had made a choice), *E* threw a switch illuminating a small red light either on the right or left side of the screen. This light indicated to the *S*s which of the two stimuli was the correct choice for that particular pair. This indicator light, along with the two stimuli, remained on for 6 sec., allowing time for a careful examination of the forms. Thus the stimuli were presented for a total of 12 sec. with an indication of the correct choice after 6 sec. The interval of time between trials was held at approximately 15 sec. This general procedure was used for all of the conditions of the experiment. The experimental setup required two *E*s. One experimental assistant operated the slide projector while another recorded the responses and their latencies on prepared data sheets.

Subjects.—The *S*s were 384 airman basic trainees at Lackland Air Force Base. Twenty-four *S*s were assigned to each of the 16 primary experimental conditions.

Each group of *S*s was given general instructions in the operation of the response keys. The *S*s were told that their task was to make as many "correct" choices as possible as indicated by the position of the red light. No mention was made of the possibility of identical segments in the contours successively presented. Thus, the task involved the discovery of the basis for a "correct" choice as well as the fixation of this response through repeated reinforcement.

Stimuli.—The stimulus forms employed in this study are shown in Fig. 1. Although six forms are illustrated as constituting each arbitrary group or class, actually 12 2 × 2-in. slides were constructed for each condition. Six of these slides reversed the right-left position of the forms as they appeared on the screen.

Figure 1 illustrates the manner in which the number of common segments was varied through four degrees. Within each of these degrees the "correct" (plus) and "incorrect" (minus) groups are indicated. The particular stimulus forms always paired together are indicated within each of the degrees by the successive rows. Thus for each degree, six pair of stimulus forms are shown. In Fig. 1 consistency of reinforcement is held constant at 100%.

The general principles followed in the construction of the forms may be described with reference to the zero common-segment condition. Each of the forms in this condition, and throughout all of the conditions of the experiment, is composed of some three-segment combination of 18 basic segments joined to form a closed contour. These basic segments were constructed by asking a group of *S*s unfamiliar with the experiment to draw any continuous irregular line they wished between two specified points. (These two points represented a distance equivalent to one-third of the circumference of a circle of given size.) From these 18 basic segments three-segment combinations were chosen to form two groups of six forms each. Certain principles were followed in the choice of segments for each of these two groups. Each of the 18 segments appears once in each group but never in the same three-segment combination. This balancing eliminates any obvious communality in one group not present in the other as well. Note also that the designation "zero common segments" is valid only with reference to the six-member groups of forms individually, and not with reference to pairs of such groups.

In the condition specified as "one common segment" each group is characterized by the possession of a common segment unique to that group. This common segment is indicated at the side of each group. Note that the common segment appears at a different compass orientation within each of the six forms. The communality or homogeneity within the group of forms is increased successively by the introduction of a second and then a third common segment. The common segments and their junctures are shown at the side of each group. With the addition of a third common segment the forms within each group differ among themselves only in degree of rotation about the transverse axis. It may be noted that as the number of common segments within each group is increased

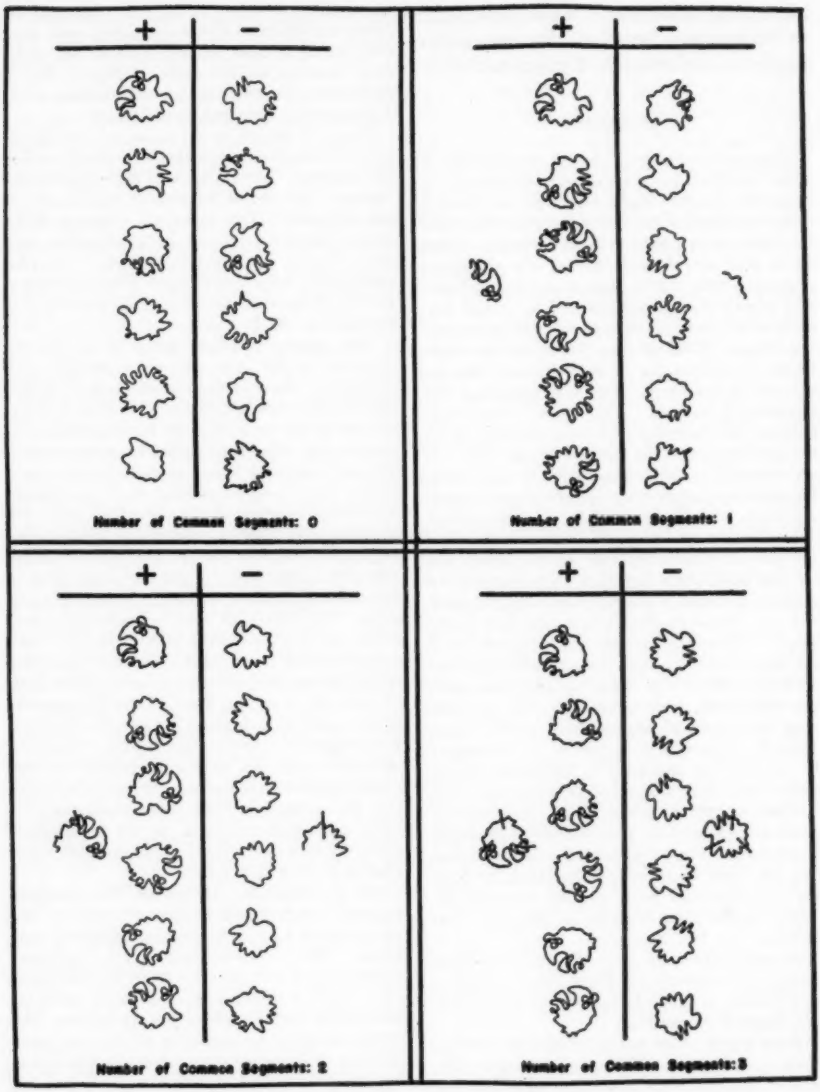

FIG. 1. Stimulus forms arranged to illustrate the increase in common segments from zero to three within paired six-member groups of forms. Plus and minus symbols are used to indicate the correct and incorrect stimuli. The common segments and their junctures are indicated at the side of each group.

TABLE 1

ERROR MEANS AND STANDARD DEVIATIONS
(N in each group = 24)

Consistency of Reinforcement	Number of Common Segments							
	0		1		2		3	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
0%	26.88	13.20	31.29	11.85	39.63	12.34	46.83	6.55
33%	26.96	10.24	27.58	10.49	28.00	12.24	38.50	8.38
67%	28.10	12.55	15.90	9.67	20.33	8.22	22.75	10.94
100%	24.42	11.48	4.92	5.38	3.63	6.59	0.90	1.05

successively, certain segments which were balanced between the two paired groups drop out altogether. This procedure must be followed if the number of common segments is to be increased without increasing the total number of segments within each form. Since this procedure of replacing "noncommon" segments with common segments was employed throughout the study, the variable represented may be thought of as a parameter.

For purposes of this study, consistency of reinforcement was varied in four degrees by interchanging the "reinforcement values" of a certain proportion of the forms from the paired groups. To produce variation in consistency of reinforcement an identical procedure was followed for each of the four similarity conditions. For 100% consistency the six forms containing the same segments were all correct; likewise the paired six forms were all incorrect. Sixty-seven per cent consistency was defined as the condition in which five out of six of the similar forms were correct and thus one out of the six was incorrect.[2] For 33% consistency the reinforcement values of two out of the six forms were interchanged; finally, with 0% consistency half of the original group of similar forms was given opposite reinforcement. (Note that per cent consistency was actually the difference between the per cent correct and the per cent incorrect.) The zero common-segment condition constitutes a control. Although there is no "group of similar forms" (necessary if the concept of consistency is to be meaningful) the reinforcement values of specified proportions of forms were interchanged in a manner identical to the other similarity conditions to test for the hypothesized convergence of the functions.

Procedure.—In terms of the primary independent variables the experiment was factorial in design, i.e., each level of one variable was tested with each level of the other variable. In all, 16 experimental conditions were involved, since the two independent variables were each manipulated in four degrees.

The testing procedure consisted of 96 slide presentations. Each of 12 slides was shown eight times, or each pair 16 times. The 12 slides appeared in a different order in eight consecutive 12-trial blocks according to a prearranged order of presentation. The sequence of positions for the correct form of each pair (the correct form was indicated by the position of a signal light) was predetermined such that the number on the right and left was equated and systematic sequences were avoided. With this arrangement the pairs of forms with interchanged reinforcement values, i.e., the "inconsistent" instances, appeared twice in each successive block of 12 trials (once on the right and once on the left). However, the sequence of trials was arranged to avoid systematic "massing" or "spacing" of these inconsistent instances.

The assignment of *Ss* to the various experimental conditions could not be completely random since four *Ss* were tested at one time. The *Ss* were assigned in such a way, however, that systematic differences due to individual difference variables were balanced among the primary conditions of the experiment.

RESULTS

Tables 1 and 2 show the mean total error and mean response latency under

[2] A possible confusion of the present design with "partial reinforcement" procedures should be clarified. The present use of "inconsistency" does not refer to a change in the reinforcement value of any form during the presentation period for a given group of *Ss*. Rather, the inconsistency lies in the fact that when the same response (choice) is made to a group of similar forms, this response for certain specific forms was *never* reinforced.

TABLE 2

LATENCY MEANS AND STANDARD DEVIATIONS (IN SECONDS)
(N in each group = 24)

Consistency of Reinforcement	Number of Common Segments							
	0		1		2		3	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
0%	2.09	0.37	2.09	0.37	1.93	0.48	2.22	0.33
33%	1.89	0.37	1.99	0.44	1.87	0.39	2.06	0.60
67%	2.14	0.47	1.52	0.42	1.36	0.50	1.41	0.43
100%	1.86	0.29	1.04	0.43	0.83	0.27	0.70	0.12

each of the experimental conditions. Essentially the same data are shown in graphic form in Fig. 2 and 3.

The close approximation of the results to those hypothesized is immediately apparent. A set of converging functions was obtained which represents the interaction between number of common segments and consistency of reinforcement. With high similarity between the forms of a class (three common segments) variations in consistency of reinforcement from 0% to 100% are seen to vary discrimination-learning performance, as measured by mean total error in 96 trials, from 46.83 (slightly above chance) to .9. This is almost the entire range of learning performance possible under the present conditions. As the similarity within the class of forms is decreased, the range of performance is seen to decrease progressively. Finally, with "no" similarity within the class of forms (zero common segments), variations in consistency of reinforcement produce what are found to be chance differences in performance error. For the latency measure the picture is somewhat less clear for the 0% consistency condition. Possibly the group testing situation and the limitation on response time to 6 sec. are responsible for the "ceiling"

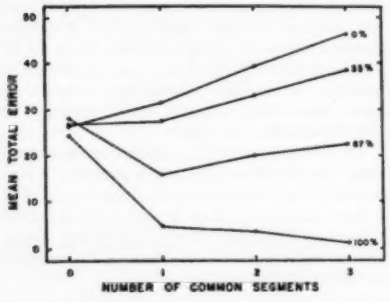

FIG. 2. Mean total errors for the 96-trial testing period as a function of number of common elements and percentage consistency of reinforcement

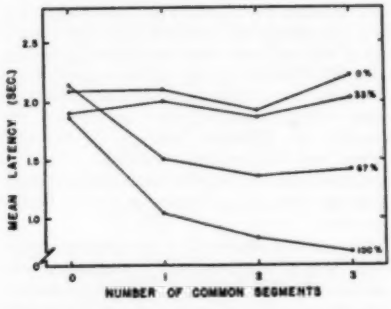

FIG. 3. Mean response latency per trial as a function of number of common elements and percentage consistency of reinforcement. The latencies for each S were averaged for the 96-trial testing period and the mean of this averaged value for each of the 24 Ss obtained.

at about 2 sec. apparent in the data of Fig. 3.

Separate analyses of variance were performed for the error and latency data. In brief, for both the error and latency analyses number of common segments, consistency of reinforcement, and the interaction of these variables were all highly significant ($p < .001$).[3]

In lieu of separate *t*-test comparisons, a single confidence interval at the .05 probability level was established for the error and latency data separately. As an estimate of the universe variance (s^2) the mean square for *S*s from the corresponding analysis of variance was used. For the error data, a difference between means of 5.76 or more was determined to be significant at the .05 confidence level. Thus, as may be seen from Table 1, with three common segments the smallest mean difference is significant at this level while with zero common segments the largest mean difference does not achieve significance. For the latency data, a difference between means of .24 was determined to be significant. From Table 2 it is seen that at the three common-segments level the smallest mean difference is significant while at the zero common-segments level two comparisons just achieve significance. The conclusion is justified, therefore, that both in the over-all analysis and for more detailed comparisons the experimental hypothesis has been substantiated unequivocally.

[3] It is evident from an examination of Tables 1 and 2 that with such a wide range of performance an appreciable correlation between the means and *SD*'s is present; therefore, the assumption of homogeniety of variance is quite likely not satisfied. In view of the extremely high *F* ratios for consistency of reinforcement (in the error analysis 133; in the latency analysis, 104; with *df* in both cases, 3/368) the question of homogeneity of variance appears to be somewhat academic.

Discussion

The effect of increasing the number of common segments within a group of stimulus forms has been shown to be a function of the proportion of these stimuli designated "correct" relative to the proportion designated "incorrect." The present study would seem to provide an experimental framework for a systematic approach to a variety of discrimination-learning and concept-formation tasks. In addition, the results would seem to have theoretical implications for both discrimination-learning and transfer-learning situations in which the communality within groups of stimuli and the consistency of reinforcement relative to these groups may be specified or varied.

On a molecular level the concept of stimulus generalization may be used to account for the results. According to the present view, the increase in the number of common segments may be assumed to have increased the amount of stimulus generalization among the stimuli. When the entire group is "correct," such generalization is undoubtedly facilitative in its effect. Such facilitation would certainly be suggested by the small number of trials required to learn the "correct" forms with three common segments and 100% consistency of reinforcement. With less than 100% consistency it appears reasonable to suppose that the trials on which some proportion of the original group of stimulus forms are presented as "incorrect" represent, in a sense, experimental extinction of the response which has been made to the majority of the stimuli. Thus, as conceived here, variation in consistency of reinforcement changes the relative weights of a facilitative and an inhibitive factor. This inhibitive factor is assumed to generalize among all of the group just as the facilitative factor does. Accordingly, the probability of a "correct" response is conceived to be, in part, a function of the algebraic summation of these opposing effects, each weighted according to its proportional representation.[4] As the proportion of

[4] It is not clear to what extent the specific concepts employed by Hull and his associates are applicable to the present discrimination-learning situation. In general, however, the present formulation appears consistent with rein-

"correct" and "incorrect" stimuli approaches one-half, the magnitude of the inhibitive factor is assumed to approach that of the facilitative. One would expect task difficulty to increase, therefore, whenever consistency of reinforcement is decreased, assuming some generalization among the stimuli in question.

The interaction between the primary variables is revealed by the fact that the amount of change in task difficulty with variation in consistency of reinforcement is itself a function of degree of similarity. One possible explanation for this interaction is in terms of an analysis of the total habit strength for each stimulus form into proportions associated with unique and common elements. The probability of a response to such a complex stimulus is conceived to be a function of some sort of summation of individual habit strengths for elements within the form. It is assumed that only the common elements produce generalization to any measurable extent. It is reasonable, therefore, that only the proportion of the total habit strength associated with common elements should be affected by variation in consistency of reinforcement. As this proportion is decreased, the effect of variation in consistency of reinforcement is decreased likewise. In the present study the proportion of unique to common elements was varied directly, since the total number of elements was held constant. It should be evident, however, that wherever there is an increase in communality or similarity among a group of stimuli, either in the laboratory or in nature, there is ordinarily a concomitant decrease in uniqueness or individuality. Even where the unique elements are not actually eliminated, they seem to become less "distinctive" as more and more common elements are added. An explanation of the interaction in these terms, therefore, would seem to have considerable generality.

Assuming the present formulation in terms of stimulus generalization adequately accounts for the results, one might expect to get a similar set of interacting functions wherever degrees of stimulus generalization may be specified or varied with concomitant variation in consistency of reinforcement. It is not at all difficult to find examples in the literature of variations in stimulus similarity with results comparable to the present study; it is difficult, however, to find cases in which stimulus similarity and a variable comparable to consistency of reinforcement were manipulated at the same time. In the discussion which follows an attempt will be made to integrate the present approach with some of the alternative formulations which have been proposed.

A number of theoretical integrations of the relationship of stimulus and response similarity to transfer effects have been formulated. The formulations of Bruce (**1**), Gibson (**4**), Osgood (**7**), Underwood (**10**), and Gagné, Baker, and Foster (**3**) are perhaps the most complete and most well known. Though not always described in such terms, to the present writer each of these formulations appears to be concerned, at least in part, with the interaction between stimulus similarity and a variable most frequently identified as response similarity. Let us examine this latter variable more closely. Obviously the *response* of *S* cannot be an independent variable manipulated directly by *E*. Nor is it likely that the writers using this term in connection with the paired-associates procedure conceive of response similarity in this manner. It would most likely be agreed that the so-called response similarity variable is actually variation in the response *characteristics* which are considered "correct" or which are otherwise encouraged. When considered in this sense, response similarity is clearly a reinforcement variable and may be manipulated as such. By changing response similarity in the transfer task, *E* can change the degree to which the original response is still reinforced, or, in a sense, the proportion of the original response which is reinforced relative to the proportion which is extinguished (assuming response similarity has been quantified).

In the present discrimination task the members of a class of similar stimuli are presented repeatedly and consistency of reinforcement is defined in terms of the proportion of the stimuli for which the same response is reinforced. In the transfer situation, on the other hand, using the paired-associates procedure, each class of similar stimuli is ordinarily represented by a single stimulus item in the original and transfer lists. If the response items are identical for corresponding stimulus items of the two lists, in effect, 100% of the original response is reinforced. If the material employed permits quantification in terms of number of common elements, the proportion of the original response which receives reinforcement in the transfer list may be decreased by degrees, i.e., response similarity may be decreased. According to the present view this variation in the proportion of the response reinforced produces, in terms of intervening process, an *effect* equivalent to variation in consistency of reinforcement as defined here.

When similarity has been defined in terms of relative proximity of stimulus values (ordered along either a physical or response-defined dimension), it is perhaps less meaningful to speak of "the proportion of the original response

---

forcement theory, i.e., Hull (**5**), and specifically with the analysis of discrimination learning in animals outlined by Spence (**8, 9**).

which is reinforced." In a recent article, Morgan and Underwood (6) proposed that the effect of varying response similarity might be accounted for in terms of the concepts of response generalization and "parasitic reinforcement." According to these authors, at the same time that a given response is reinforced in the first list there is, in effect, "parasitic reinforcement" of similar responses. It is assumed that this reinforcement decreases in degree as the generalized response is less similar to the one actually reinforced. They reason that one would expect an increase in response similarity to increase the amount of associative facilitation, since some of the response items may already have corresponding response tendencies at a "supra-threshold" level. The explanation offered by Morgan and Underwood is essentially the same as that suggested here in terms of consistency of reinforcement. According to the present formulation, however, when the original response is reinforced, in a sense, connections are strengthened between "elements" (or dimensional values) of both stimulus and response. To the extent that the transfer task duplicates these stimulus "elements," the corresponding response "elements" will at least have a greater-than-chance probability of being evoked. The degree of associative facilitation or inhibition in the transfer task, therefore, is considered to be a function of the proportion of response "elements" which are reinforced relative to the proportion extinguished.

As further evidence for the view that response similarity and consistency of reinforcement are functionally equivalent, it may be shown, for example, that the "transfer and retroaction surface" proposed by Osgood (7) may be translated directly into the form illustrated by Fig. 2 in the present study. To make this translation it is only necessary to substitute consistency of reinforcement for response similarity and to take the point of convergence of the functions at zero common segments as the base line from which negative and positive transfer effects are determined.

Since the implications of the present study go far beyond the limits of the discrimination learning task investigated here, the need is indicated for a separate theoretical discussion applying the concept of consistency of reinforcement both to original-learning and transfer-learning situations. The present study may be thought of as the empirical basis for this proposed theoretical formulation, the outlines of which have been sketched briefly.

Summary

1. Using a discrimination-learning task with stimuli consisting of closed irregular contours, this study tested the hypothesis that discrimination-learning performance is a joint function of (*a*) the number of common stimulus elements within a group of forms, and (*b*) the proportion of these similar forms for which the response of "choice" is reinforced or indicated as "correct." According to the hypothesis, with high consistency of reinforcement increasing the number of common elements decreases the difficulty of the discrimination task progressively. As the consistency of reinforcement is reduced, however, the function describing decreasing difficulty changes slope progressively and is finally replaced by a function describing increasing difficulty. Thus, in effect, an interaction between the two independent variables was hypothesized.

2. Each closed irregular contour was composed of three joined contour segments. Three-segment combinations of 18 basic segments were so arranged that the number of identical contour segments could be varied from zero to three within different groups of stimuli. In such groups, variation in four degrees was also introduced in the percentage of forms designated "correct" relative to the percentage designated "incorrect." This latter manipulation defines variation in consistency of reinforcement.

3. The discrimination task employed was that of learning to choose on each trial the "correct" form in successively presented pairs of such forms. The forms in each pair were individually from each of two six-member groups of forms within which the number of common segments was varied.

4. The procedure involved testing the discrimination performance for a 96-trial testing period, with each of the 16 possible combinations formed by the four possible degrees of varia-

tion of the two independent variables. Twenty-four airman basic trainees each were assigned randomly to each of the 16 experimental conditions.

5. The results constitute an experimental verification of the hypothesized interaction between number of common segments and consistency of reinforcement at a high level of confidence.

6. A theoretical formulation of the results in terms of stimulus generalization is outlined. The apparent equivalence of consistency of reinforcement and "response similarity" is discussed.

(Received July 11, 1952)

References

1. Bruce, R. W. Conditions of transfer of training. *J. exp. Psychol.*, 1933, **16**, 343–361.
2. French, R. S. A comparison of rote and discrimination learning with stimulus homogeneity variant. Unpublished Ph.D. dissertation, Univer. of Wisconsin, 1951.
3. Gagné, R. M., Baker, K. E., & Foster, H. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67–79.
4. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
5. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
6. Morgan, R. L., & Underwood, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40**, 592–603.
7. Osgood, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.
8. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
9. Spence, K. W. The nature of the response in discrimination learning. *Psychol. Rev.*, 1952, **59**, 89–93.
10. Underwood, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949.

Journal of Experimental Psychology
Vol. 45, No. 1, 1953

STIMULUS GENERALIZATION AS A FUNCTION OF LEVEL OF EXPERIMENTALLY INDUCED ANXIETY[1]

GERALD ROSENBAUM

Wayne University

In recent years, the term *anxiety* has been used in theoretical interpretations of behavioral phenomena to mean a hypothetical state or response accompanying or resulting from the anticipation of noxious stimulation. According to this formulation, first proposed by Mowrer (**15**), an implicit anxiety response becomes conditioned to previously neutral cues which have been associated with painful stimulation. The conditioning of this internal response is assumed to provide an acquired motivational state in that the presence of anxiety cues leads to more vigorous activity and escape from these cues serves to reinforce the learning of new responses. Animal studies have been conducted by Miller (**13**), Brown and Jacobs (**5**), and Amsel (**1**), among others, to show that anxiety does exhibit both the reinforcing and energizing properties of a drive.

The energizing function of drives is embodied in Hull's theory (**10**) by the assumption that relevant and irrelevant drives combine to multiply the habit strength of stimulus-response connections. Miller has utilized this postulate in his theoretical analyses of conflict (**12**) and displacement (**14**) to predict that variations in drive level will produce concomitant changes in the height of gradients of stimulus generalization. Since previous studies by Brown (**3**) and Rosenbaum (**16**) indicate that increases in the hunger drive result in a general elevation of performance measures of stimulus generalization, similar results might be predicted for increases in the acquired anxiety drive.

The present experiment was designed to investigate the hypothesis that stimulus generalization is a positive function of level of experimentally induced anxiety. More specifically, the prediction was made, in accord with Hull's postulates, that higher levels of anxiety would produce a greater strengthening of the response tendency to a conditioned stimulus and to stimuli differing from it along a given dimension. Performance measures of stimulus generalization would be expected to reflect this heightened gradient of excitatory strength by an elevation in the gradients of response amplitude and frequency and by a lowering of the gradient of response latency. Although increases in drive should steepen as well as elevate the theoretical generalization gradient of excitatory strength, no precise prediction of the change in the slopes of the gradients of amplitude, frequency, and latency can be made from the functional relationships Hull postulates between these response measures and excitatory strength without specifying the prevailing habit strengths and amount of drive increase. Verification of the above hypotheses would

[1] This article is based on a dissertation submitted to the faculty of the department of psychology of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express his gratitude to Professors J. S. Brown and I. E. Farber for their guidance and to the staff of the Veterans Administration Hospital, Coatesville, Pennsylvania, for their cooperation during the course of this study. An abstract of this paper was presented at the 1950 meeting of the Midwestern Psychological Association.

extend the correspondence between the notion of anxiety as an acquired drive and Hull's general drive concept, as well as elaborating the relationship between different levels of anxiety and certain performance measures of stimulus generalization.

In order to test these hypotheses, the following general procedure was employed. Human *Ss* were trained to make a simple motor reaction as quickly as possible upon the presentation of a rectangularly shaped visual figure (conditioned stimulus). Other rectangles (generalized stimuli), which differed from the training stimulus in height and to which *S* was prohibited from responding, were then presented. The frequency, amplitude, and latency of "false" motor responses to these generalized stimuli were used as measures of stimulus generalization. Three levels of anxiety were experimentally induced by training *Ss* to anticipate either a strong shock, a weak shock, or the sound of a buzzer when they responded slowly. It was assumed that *Ss* trained with more noxious stimuli would develop greater anxiety and exhibit a stronger tendency to respond to all stimuli presented.

Method

Experimental design.—The experiment was conducted in two parts, one with psychiatric patients, the other with normal college students. In the first part, 72 male psychiatric patients from the Veterans Administration Hospital at Coatesville, Pennsylvania were divided into three groups: (*a*) strong-shock, (*b*) weak-shock, and (*c*) buzzer. Each group consisted of 24 *Ss* matched for psychiatric diagnosis. The second part of the experiment was an exact replication of this design, using 72 male students from the introductory psychology course at the State University of Iowa. The replication was introduced in order to determine whether the results initially obtained with pathological *Ss* could be duplicated with normal groups.

Apparatus.—The two major functional elements of the apparatus were the stimulus-presentation unit and the response key used by *S*. The stimulus-presentation device was constructed in such a way that any one of four rectangular patches of light could be presented by *E*. It consisted of a lamp box 8 in. high, 6 in. wide, and 11 in. long, into one end of which was fixed an opal-flash glass panel (4.5 × 6.5 in.). A 60-w. lamp was mounted inside at the opposite end of the box. An opaque panel, into which four rectangular slots had been cut, could be slid back and forth behind the glass screen. When one of the rectangular slots was at the center of the box, light would pass through it to the screen, forming an illuminated image of the slot. The four vertically oriented rectangular images were .5 in. wide and varied in height from 2.25 to 3 in. in steps of .25 in. A 13 × 28-in. plywood screen, painted black and containing an aperture for the glass screen, was located in front of the lamp box to conceal *E*'s manipulation of the sliding panel from *S*.

The response key consisted of a brass slider connected into an endless loop of string that passed in front of the lamp box and around to a movement-recording system at the rear. The slider was mounted on the same baseboard as the lamp box and located in front of the plywood screen. The *Ss* were instructed to grasp a knob (1 in. long and .5 in. in diameter) at the center of the slider and to make a simple movement to the right in the horizontal plane. The extent of the required movement was indicated by a white arrow painted on the response panel 7 in. from the starting point. The maximum possible movement of the slider was 9.25 in. The slider was connected through the endless loop of string to a wooden pulley in such a way that any movement of the slider caused the pulley to rotate. Two pointers were mounted on the axle of the pulley, one being rigidly attached to the axle, the other not. The second pointer would move only with the first and remain fixed at the termination of a movement, the first returning to the starting position with the slider. The amplitude of each movement could thus be read directly from the position of the second pointer on a calibrated scale.

The latency of each response was recorded to the nearest .01 sec. by a Standard Electric timer. This timer and the 60-w. lamp in the stimulus-presentation unit were activated simultaneously by a doorbell-type switch in a holding-relay circuit, so that the clock started at the same instant that the rectangular image was presented. Movements of the slider as small as .1 in. were sufficient to break the circuit to the clock and the lamp. The holding relay prevented reactivation of the timer and the lamp when the slider was returned to the resting position. If *S* did not respond within 5 sec., *E* could break the circuit by means of a separate switch.

The shocks were generated by a Harvard inductorium energized by 1.5 v., and were administered through two copper electrodes fastened to a leather wrist band. The movable secondary coil of the inductorium was set at 5.0 cm. for the strong-shock condition and at 6.5 cm. for the weak-shock. Electrode paste was used to reduce variations in resistance. The sound of the inductorium when activated served as the buzzer.

Preliminary training and screening.—The *S* was seated at the apparatus so that his right arm rested on an adjacent table and the top of the glass screen was at eye level. For *S*s tested under shock conditions, the wrist band was then strapped on to the left wrist. All *S*s were shown the required movements of the slider and given several practice trials paired with the training stimulus. In order to screen *S*s incapable of performing the experimental task, the four stimuli were presented in a systematic order and *S* was tested for his ability to discriminate between them. If any of the other rectangles were designated as the training stimulus more than three times in the course of 20 trials, *S* was eliminated from the experiment. Although none of the normal *S*s failed to meet this criterion, six psychiatric patients were eliminated.

Training and testing for generalization.—All *S*s were instructed to move the slider as rapidly as possible when the training stimulus was presented and were prohibited from responding to any of the other three stimuli which they were told would occasionally appear. Half of the *S*s in each anxiety group were trained to respond to the 3-in. rectangle and the other half were trained with the 2.25-in. rectangle in order to control for differential effects of the extensity of the training stimulus upon the generalization gradients.

Following the reading of instructions, *S*s were given an initial series of 20 conditioning trials with the training stimulus. The training trials were followed immediately by the test series, which consisted of five test trials on each of the three generalized stimuli in addition to the training stimulus. These 20 test trials were interspersed among 60 additional training trials, so that each test trial was followed by three presentations of the training stimulus. The training stimulus thus appeared 65 times during the test series, but was recorded as a test trial only on the five occasions on which it was preceded and followed by three presentations of the same stimulus. Interspersed training trials were employed to reduce extinction effects which might have been produced by the interpolation of generalized stimuli. Twelve different orders of presentation of the test stimuli were used in each group to minimize the effects of serial position.

The three levels of experimentally induced anxiety were effected by varying both the noxious stimulation and the instructions. Accordingly, *S*s in each of the groups were told that if they responded too slowly they would receive the following treatments:

Strong-shock group—receive a painful shock
Weak-shock group—feel a slight skin stimulation
Buzzer group—hear a buzzer sound

Shocks or buzzes were administered once during the preliminary instructions and on the sixth and eighteenth training trials, regardless of *S*'s reaction latency, in order to maintain the particular anxiety level involved.

Results

The major results of the experiment obtained from measurements of the amplitude and frequency of responses are presented in Tables 1 and 2. These tables show the mean response tendency to each of the four stimuli for the normal and psychiatric *S*s under the three experimental condi-

TABLE 1

Mean Amplitude of Response in Inches to the Training and Generalized Stimuli

Group	Normal *S*s					Psychiatric *S*s				
	Rectangles				Mean	Rectangles				Mean
	0	1	2	3		0	1	2	3	
Strong shock	7.12	4.20	3.80	3.62	4.68	7.03	3.67	2.36	2.25	3.83
Weak shock	6.75	2.72	1.77	1.17	3.10	5.81	2.45	1.78	1.75	2.95
Buzzer	6.80	2.39	1.17	1.47	2.96	7.20	1.94	1.39	0.86	2.85
All *S*s	7.07	3.10	2.25	2.09		6.68	2.69	1 84	1.62	

TABLE 2

MEAN PERCENTAGE FREQUENCY OF RESPONSES TO THE TRAINING AND GENERALIZED STIMULI

Group	Normal *Ss*					Psychiatric *Ss*				
	Rectangles				Mean	Rectangles				Mean
	0	1	2	3		0	1	2	3	
Strong shock	90.8	58.3	57.5	54.2	65.2	91.7	54.2	45.0	40.8	57.9
Weak shock	83.3	37.5	26.7	24.2	42.9	77.5	36.7	35.0	31.7	45.2
Buzzer	84.2	33.3	20.0	25.8	40.8	91.7	34.2	30.0	21.7	44.4
All *Ss*	86.2	43.0	34.8	34.8		87.0	41.6	36.6	31.4	

tions. The CS is designated as 0, regardless of whether it was the largest or smallest rectangle, since comparisons of the gradients obtained revealed no essential differences attributable to the extensity of the CS employed; the other rectangles are designated as 1, 2, and 3 in order of decreasing similarity to the CS. Since the latency data were not representative of the results obtained from the other two measures, they are considered separately.

Generalization gradients.—In order to test the hypothesis that heightened anxiety tends to elevate gradients of generalization, it is first necessary to demonstrate that empirical generalization gradients were obtained with the present procedure. Inspection of Tables 1 and 2 reveals the presence of fairly continuous decrements in percentages and amplitude from the CS to the least similar rectangle in the series. This trend is apparent in all of the groups presented in the tables.

TABLE 3

ANALYSIS OF VARIANCE OF AMPLITUDE DATA FOR NORMAL *Ss*

Source of Variation	*df*	Mean Square	*F*
Anxiety levels	2	87.781	8.197*
Ss in same group	69	10.709	
Stimuli	3	364.595	187.926*
Anxiety × stimuli	6	7.855	4.049*
Pooled *Ss* × stimuli	207	1.940	
Total	287		

* Significant beyond the 1% level of confidence.

The data upon which these tables are based were subjected to analyses of variance, a sample of which is presented as Table 3. Since the same *Ss* provided the data for the response measures to the different stimuli, the procedure recommended by Kogan (11) and Edwards (6) for analysis of variance involving repeated measurements on the same *Ss* was utilized. This procedure makes use of two separate error estimates, so that the variance produced by the uncorrelated subjects in the three anxiety groups is evaluated by using an error term free of correlation, while the other major sources of variance which involve the repeated performance of the same subjects are evaluated by using an error term which takes the correlation within *Ss* into consideration. It may be noted from Table 3 that, for normal *Ss*, the variance in amplitude attributable to the four stimuli is highly significant. This finding is clearly in accord with the evidence for generalization derived from examination of the tabular data.

A graphic plot of the over-all gradient of response amplitude, obtained by pooling the data from the three normal groups, is shown as the lower solid line on Fig. 1. Comparable results were obtained from an analysis

of the patients' amplitude data and the resulting gradient has been plotted as the lower dashed line on Fig. 1. The upper curves on Fig. 1 represent the over-all gradients for both groups when only the positive responses are employed in computing the mean amplitudes, zero amplitudes (corresponding to failures to react) being excluded from the computations. It may be seen that both sets of empirical curves have a highly similar negatively accelerated form characteristic of generalization gradients. The consistently lower gradient of response amplitude for the psychiatric group may be due to the large percentage of schizophrenic patients in that group, since a generally reduced level of responsiveness is characteristic of such individuals.

Significant values of t obtained between the means of Stimuli 1 and 2 and between Stimuli 1 and 3 on both gradients revealed that the significant amount of variation between stimuli,

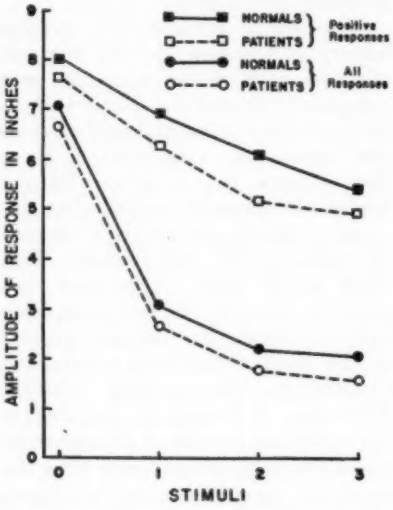

FIG. 1. Empirical gradients of response amplitude for all normal and psychiatric *Ss*

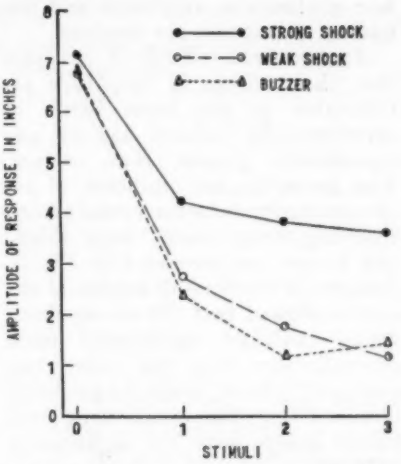

FIG. 2. Gradients of response amplitude for the three levels of experimentally induced anxiety in normal *Ss*

noted in Table 3, could not be attributed solely to the difference in response amplitude between the training stimulus and the prohibited stimuli. It may also be noted that the presence of clear-cut gradients of amplitude when zero amplitudes are excluded, precludes the possibility that the results obtained for the amplitude measure might have been an artifact of the frequency gradients.

Examination of Table 2 shows that a pooling of the percentage frequency of responses for all *Ss* yields gradients of frequency which closely approximate the negatively accelerated curves plotted from the amplitude data. Analysis of variance and t tests of these data yielded results parallel to those for amplitude. The results of these analyses indicate that when *Ss* have been trained to make a motor reaction to a particular visual stimulus and are then tested on other stimuli, to which they have been instructed not to respond, generaliza-

tion gradients in amplitude and frequency of response are obtained.

Anxiety levels.—Table 3 indicates that the variance in amplitude attributable to the three levels of experimentally induced anxiety was significantly greater than chance. The generalization gradients of response amplitude for the normal groups receiving strong shock, weak shock, and buzzer are presented in Fig. 2. Analysis of the over-all heights of the curves showed that the strong-shock group exhibited significantly more generalization than the other two groups (1% level), while the gradients obtained for the weak-shock and buzzer groups were not significantly different.

Analysis of the amplitude data obtained from the psychiatric patients yielded results comparable to those obtained from the normal *S*s. Inspection of Table 1 reveals that the over-all gradient for the strong-shock group was again markedly higher than those obtained for the other two groups.

The effects of induced anxiety upon the gradients of frequency may be seen in Table 2. Statistical treatments of these data indicated that strong shock produced a significant elevation of the gradient of frequency over the gradients obtained from the other two conditions in both the normal and psychiatric groups.

Examination of Fig. 2 suggests that strong shock, in addition to elevating the generalization gradient, tends to flatten it. A statistical test of this effect involves an evaluation of the differential effects of the three levels of experimentally induced anxiety at the four points along the stimulus dimension. This evaluation is accomplished in the analysis of variance by the *F* ratio attributable to the interaction between anxiety level and stimuli, shown in Table 3 to be significant at better than the 1% level. Equivalent flattening of the gradient with increases in anxiety level was also found for the psychiatric *S*s and was present in the frequency data of both groups.

This finding is in accord with the common clinical observation that high levels of anxiety tend to impair discrimination, since accuracy of discrimination between stimuli should be a function of the steepness of the generalization gradient rather than its absolute height. It should be noted, however, that the present findings may have resulted from the artificial maximum imposed on the measures of response amplitude and frequency used in this experiment. In terms of Hull's theoretical assumptions, some flattening of the empirical generalization gradients should accompany drive increases whenever the observed response to the conditioned stimulus is near its physical maximum, as increases in excitatory strength would then tend to elevate only the responses to generalized stimuli.

Measures of latency.—The median latency of response for the five test trials on each of the rectangles was computed for every *S*. Table 4 shows means of these individual median reaction times. An inspection of these data reveals that reaction latency tended to decrease, rather than increase, with greater disparity between the training and test stimuli. This trend would seem to indicate that response strength, as estimated from latency, tends to become greater as the similarity of the generalized stimulus to the training stimulus decreases. A finding to this effect is obviously the reverse of the generalization phenomenon obtained from the amplitude and frequency measures. However, since latency and frequency are positively related (short mean

TABLE 4

MEAN OF MEDIAN LATENCIES OF RESPONSE IN SECONDS TO THE TRAINING AND GENERALIZED STIMULI

Group	Normal *Ss*					Psychiatric *Ss*				
	Rectangles				Mean	Rectangles				Mean
	0	1	2	3		0	1	2	3	
Strong shock	.39	.31	.34	.30	.34	.55	.42	.45	.29	.43
Weak shock	.56	.46	.24	.24	.38	.54	.27	.32	.25	.34
Buzzer	.49	.42	.34	.29	.38	.61	.60	.46	.22	.47
All *Ss*	.48	.40	.31	.28		.57	.43	.41	.25	

latencies being associated with low mean frequencies), it is apparent that the number of measures used in the computation of the individual latency scores decreased with the gradient of frequency. This means, in effect, that the number of indeterminately long latencies (corresponding to failure to react) increased with the distance along the stimulus continuum from the training stimulus. Had it been possible to enter these long latencies into the computations, the direction of the resulting gradients would quite probably have been the same as those obtained for amplitude and frequency. This explanation obviously does not account for the fact that the latencies which did enter into the computations yielded a gradient which was the reverse of generalization, rather than assuming some other form.

Since generalization gradients were not obtained from the latency data, it was not possible to use this measure in testing the hypothesis that amount of generalization is positively related to anxiety level. However, the effect of anxiety level upon over-all reaction latency was considered by computing the median latency to all test stimuli for each *S* and comparing the three groups. The results of this analysis showed that the mean latency for the strong-shock *Ss* was significantly shorter than the mean for weak shock at the 5% level in the normal groups. No other significant effects were observed as these data were extremely variable, particularly in the psychiatric groups.

DISCUSSION

It has been shown in the present study that when *Ss* have been trained to make a voluntary motor response to a particular rectangle and then tested on other rectangles, the strength of the response tendency to the test stimulus varies inversely with the distance, along a size dimension, between that stimulus and the training stimulus. Voluntary responses have previously been employed in studies of generalization by Gibson (**7**) and by Brown, Bilodeau, and Baron (**4**). Both of these studies have reported negatively accelerated gradients of response frequency which were quite similar to the gradients of frequency and amplitude obtained in the present investigation. The empirical generalization curves obtained in all three studies have shown a close correspondence with those obtained by Hovland (**9**) and Bass and Hull (**2**) for the conditioned galvanic skin response. This correspondence lends support to the assumption that the

generalization process is similar for both voluntary and conditioned responses.

In all three studies of voluntary responses, however, the latency gradients were in the opposite direction of those obtained from generalization of conditioned responses. This finding illustrates one essential difference between the two generalization phenomena, namely that generalization gradients obtained in studies employing voluntary responses result from the combined action of long-latency, verbally mediated responses and short-latency, relatively unmediated responses. The inverted gradient of latencies is produced by the presence of mediated reactions to stimuli similar to the training stimulus resulting in long mean latencies, while mediated response tendencies toward less similar stimuli result in failures to react and short mean latencies. Since the tendency to label generalized stimuli as "incorrect" is produced by the instructions prohibiting responses to these stimuli, the mediated tendency to inhibit reactions should be constant for all such stimuli. Differences in the frequency of unmediated responses to the stimuli along the generalization continuum would then be a function of primary generalization of differential amounts of excitatory strength, produced by the practice with the training stimulus. However, a greater frequency of long-latency, mediated reactions to stimuli similar to the training stimulus should also be expected, since any implicit responses which were present when a positive reaction was made to the training stimulus should also generalize and serve to mediate reactions to similar stimuli.

In effect, this formulation suggests that the generalization of voluntary responses involves not only a decreasing primary gradient of response strength toward less similar stimuli, but, in addition, a parallel gradient of mediated response tendencies resulting from the secondary generalization of responses to the perceptual and verbal response-produced cues associated with the stimuli along the generalization dimension. A prediction which might be generated by this formulation is that a more conventional generalization gradient of latencies would be obtained if the tendency to mediate voluntary reactions were minimized by a systematic emphasis on speed.

In general, the predictions derived from the relationship of the drive concept of anxiety to stimulus generalization were substantiated in the present experiment. Although the noxious stimuli (shocks or buzzer) were presented only three times, relatively early in the course of a hundred trials, a marked difference in the level of responsiveness to similar stimuli was noted between the strong-shock and the other conditions. The difference obtained between the strong- and weak-shock conditions suggests that the rate of rise of the conditioned anxiety response is more rapid in situations involving strong shock. Since the number of anxiety conditioning trials was held constant, the anxiety response should have reached a higher maximum in the strong-shock condition. The fact that no difference was obtained between the weak-shock and buzzer groups suggests that the maximum reached by the anxiety response under these conditions was either zero or a small positive value. These results are in accord with the findings of Gwinn (8) that the strength of the fear response is a function of the intensity of the noxious stimulus.

The present study has shown that the amount of stimulus generalization, as measured by the over-all heights of the frequency and amplitude gradients, is positively related to the degree of experimentally induced anxiety. Since similar elevations of generalization gradients have been observed with other drive states, the heightened responsiveness of anxious individuals may be regarded as one instance of the more general effects of strong drive states upon the generalization of response tendencies.

Summary

The present experiment was concerned with the effects of three levels of experimentally induced anxiety upon the generalization of a voluntary response to visual stimuli. It was hypothesized that higher levels of anxiety would produce corresponding elevations of the generalization gradients.

Seventy-two male psychiatric patients and an equal number of college students were each divided into three anxiety levels, effected by the use of two different intensities of shock and a buzzer. All *Ss* were trained to make a motor reaction to a visual stimulus and then tested on other stimuli, differing in height from the training stimulus. The frequency, amplitude, and latency of the responses to these generalized stimuli were employed as measures of generalization.

The results showed that the shapes of the generalization gradients for fre-

quency and amplitude were highly similar in both the normal and psychiatric groups. Latency of response yielded the inverse of a decreasing gradient of response strength. However, it is suggested that a generalization gradient would be obtained with this measure if indeterminately long latencies were considered.

Strong shock resulted in significantly higher generalization gradients than either weak shock or buzzer with both normal and psychiatric *S*s. No differences were found between the effects of a weak shock and a buzzer. The gradient was flatter under the strong-shock condition than under the other two conditions.

It was concluded that experimentally induced anxiety shows the energizing function of a drive in elevating generalization gradients. The discrepancy between the latency data and the results obtained with other measures is discussed and interpreted in terms of the effects of mediated processes upon generalization gradients obtained with voluntary responses.

(Received July 16, 1952)

References

1. Amsel, A. The combination of a primary appetitional need with primary and secondary emotionally derived needs. *J. exp. Psychol.*, 1950, **40**, 1–14.
2. Bass, M. J., & Hull, C. L. The irradiation of a tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, **17**, 47–65.
3. Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.
4. Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.
5. Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
6. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
7. Gibson, E. J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, **24**, 237–253.
8. Gwinn, G. T. The effects of punishment on acts motivated by fear. *J. exp. Psychol.*, 1949, **39**, 260–269.
9. Hovland, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
10. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
11. Kogan, L. S. Analysis of variance—repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.
12. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.
13. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
14. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
15. Mowrer, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–565.
16. Rosenbaum, G. Temporal gradients of response strength with two levels of motivation. *J. exp. Psychol.*, 1951, **41**, 261–267.

Journal of Experimental Psychology
Vol. 45, No. 1, 1953

THE ROLE OF INSTRUCTION IN EXPERIMENTAL SPACE PERCEPTION [1]

CECIL W. MANN AND RANDOLPH O. BORING [2]

Tulane University

Perceptions and the descriptions of perceptions are selective processes depending upon a number of factors, some of which are mental set, training, experience, and purpose. The perception of a landscape by an artist, a poet, a forest ranger, and a house builder will be different, and so will be their descriptions. The testimony given in a court of law by a witness answering specific questions will probably differ from that given when he describes events in response to the unstructured request to tell what happened.

In evaluating perceptual experiments it is often necessary to know to what extent the results have been influenced by selective factors. In order to clarify this situation, it is proposed to group perceptual experiments into two types. Type X will include those experiments in which *S* performs and *E* observes the performance. In the Rorschach experiment, for instance, the directions to *S* are designed to keep him at the level of a naive observer. It is *E* who is the observer for purposes of this experiment, and it is he who needs sophistication by training and experience to the point where he can make significant observations. Similarly, in a card-sorting experiment, *S* needs to be sophisticated merely to the point at which he knows the rules of the game; *E* makes the observations and must be trained in the selection of relevant details for observation.

Type Y experiments are those in which *S* is the observer as well as the performer. He is faced with the task of observing the effect of the variation of some condition upon his perceptual processes. In experiments of this type, *E* retires to the position of mechanic who sets the conditions of the equipment or clerk who records. In measuring auditory threshold, for instance, *E* manipulates frequencies and intensities, *S* makes the observations. The *E* has no subjective check on what *S* perceives; he cannot listen for *S*. His check will be objective: the degree of consistency of reports by the observer upon replication of the conditions.

It is in the latter experiments that the selective factors operating on *S* become important. If the effect to be observed is complicated, ill-defined, or confused by other or similar effects, the consistency of the reports may be greatly reduced. Reports of the oculogyral effects illustrate this point. Very few naive *S*s report the observation of all effects during and following rotation prior to the fifth or sixth rotational trial, and some *S*s require ten or twelve trials before they are able to make full reports. The reports of the duration of the effects vary considerably during the first few trials, and thereafter increase in consistency with their selective perception.

[1] This report is one of a series of investigations conducted jointly with the U. S. Naval School of Aviation Medicine under Contract N7-onr-434 T.O.1, with the Office of Naval Research, in cooperation with the Bureau of Medicine and Surgery, and the Tulane University of Louisiana.

[2] The senior author is responsible for the article in its present form. Randolph O. Boring modified the equipment and collected the data.

The effect of selective perception is illustrated by a series of experiments designed to determine the precision of the judgment of the vertical under varying conditions of visual and body tilt. It has been variously argued that under conflicting visual-proprioceptive conditions the determinants of the perceived vertical will be the visual factors (**5, 6**) or the proprioceptive factors (**3, 4**). In an attempt to resolve this confusion, Gibson (**2**) has recently suggested that more emphasis be placed upon the joint interaction of the factors in the visual-proprioceptive complex. He argues that when the main lines of the retinal field are in alignment with the direction of the gravitational force, the visual factors will interact with the compensatory mechanism of the statocysts to produce an invariant resultant. Under this condition of *consistent* variables the perception of the vertical will be precise. This condition is produced whether the body is upright or tilted, for the tilt of the body does not change the direction of gravity, albeit the force does pass through the body at an unusual angle. When the retinal framework is not in alignment with the gravitational force —because of the rotation of the visual field by mechanical or optical methods, or by the addition of a centrifugal component to that of gravity—the condition is *discrepant*, and the visual-proprioceptive variables are in conflict. The judgment of the vertical under this condition will be less precise.

Procedure

In the design of this experiment the Tulane lateral-tilt chair was used with the addition of a small visual framework placed at eye level at a distance of 5 ft. in front of *S*. Two frameworks were used: (*a*) a luminescent straight line ¼ in. wide and 14 in. in length; and (*b*) a luminescent window pattern 12 × 14 in. bisected vertically and horizontally. The chair and framework could be tilted independently to produce consistent or discrepant conditions. The *S* was required to set a selsyn-controlled target rod, ¼ × 12 in. to the direction of the gravitational vertical while maintained by *E* in the preset consistent or discrepant condition. Three angles of inclination were used for chair and framework, 30°L, 0°, and 30°R, and *S*s made five judgments in each of the nine conditions, arranged in random order. The target rod was set at 30°R and 30°L, an equal number of times in random order. The trials were all made in the absence of visual cues. The *S* closed his eyes while *E* set the conditions, and, upon instructions, opened his eyes and made his judgment by adjusting the target rod to the gravitational vertical. The *S* had no control of the tilt of the chair or the framework, but could move the target rod in either direction by the use of a small switch close to his right hand. The framework and the chair could be set by *E* to an accuracy of 1°; the target rod was geared to a recorder and could be read to an accuracy of .1°.

It will be noted that under some conditions *S* has to make a choice of one of three verticals. For instance, when his body is tilted at 30°L and the visual framework at 30°R, he may (*a*) accept the tilt of his body and set the target rod in alignment with his body or in a position that he believes will compensate for the tilt, (*b*) accept the visual framework and align the target rod thereto, or (*c*) attempt to set the target rod in alignment with the upright walls of the room housing the tilt chair. Furthermore, in any series of trials, unless he is constantly aware of his criterion of verticality, he may change his choice from one to another. One would predict that naive *S*s with relatively unstructured directions would be less precise and more variable than *S*s sophisticated with respect to the criterion. In order to test this hypothesis, the *S*s—undergraduates ranging from 18 to 20 yr.—were divided into two groups with four in each group.

Naive group.—Each member of the naive group was brought into the darkened room, and was told that the problem was to set the target rod to the vertical. He was shown how to control the movement of the target rod. No instructions were given with respect to what constituted verticality, and if he asked questions, he was told to use his own judgment in setting the target rod.

Sophisticated group.—Each member of this group was brought into the lighted room. He was seated in the chair and shown how to control the target rod. He was told that his problem was to align the target rod to the gravitational vertical, that is, either parallel to the walls of the room, or at right angles to the floor. He was then given a few trials in the lighted room in order to make sure that he was aware of the

TABLE 1

MEAN AVERAGE ERRORS IN THE JUDGMENT OF THE VERTICAL FOR NAIVE AND SOPHISTICATED GROUPS UNDER DIFFERENT CONDITIONS OF TILT OF VISUAL FRAMEWORKS I AND II AND BODY TILT

Chair Position	Group	Frame 30°L		Frame 0°		Frame 30°R	
		I	II	I	II	I	II
30°L	Soph.	3.0	2.4	3.4	4.6	2.0	1.8
	Naive	17.0	16.3	11.8	12.0	4.4	2.4
0°	Soph.	1.7	3.3	1.4	1.8	1.4	2.0
	Naive	7.9	7.8	2.8	2.3	8.1	10.3
30°R	Soph.	3.4	1.6	3.5	2.5	3.9	5.9
	Naive	4.7	5.4	13.6	12.0	20.0	21.4

criterion of verticality. He was encouraged to ask questions, and insofar as these were pertinent to the criterion of verticality, they were answered.

Every effort was made in giving directions to ensure that the naive group maintained its own criterion of verticality and that the sophisticated group accepted that of the experimenter.

RESULTS

Mean gross average and mean constant errors are shown for each group and for each visual framework in Tables 1 and 2, respectively. It is not proposed to discuss here the almost negligible influence of the visual framework upon the judgments of the sophisticated group since these results have been fully dealt with elsewhere (**1, 4**). More important in this paper is a comparison of the results of the naive and the sophisticated groups. It will be seen that in every comparison, even in those where the stimulus variables are consistent, the errors made by the naive group are larger than those of the sophisticated group.

An analysis of the constant and average errors of the sophisticated group shows homogeneity of variance and differences between conditions and subjects that are not significant. An analysis of the results of the naive group would not appear to be useful because of the great variability. For instance, two *Ss* made gross average

TABLE 2

MEAN CONSTANT ERRORS IN THE JUDGMENT OF THE VERTICAL FOR NAIVE AND SOPHISTICATED GROUPS UNDER DIFFERENT CONDITIONS OF TILT OF VISUAL FRAMEWORKS I AND II AND BODY TILT

Chair Position	Group	Frame 30°L		Frame 0°		Frame 30°R	
		I	II	I	II	I	II
30°L	Soph.	1.0L	2.0L	3.5L	4.4L	0.9L	1.5L
	Naive	14.3L	8.3L	8.4L	11.4L	0.5L	1.3L
0°	Soph.	1.6L	2.8L	1.0R	0.4L	0.7R	1.2R
	Naive	5.7L	6.3L	1.9R	1.8R	7.6R	9.2R
30°R	Soph.	3.4R	0.5R	1.6R	0.9R	3.1R	4.4R
	Naive	3.7R	2.6L	13.6L	10.3L	18.3R	20.3R

errors of more than 50°, while under the same condition the other two *Ss* made gross average errors of less than 5°. Under two-thirds of the conditions represented in Tables 1 and 2, mean values for naive *Ss* are misleading because of the wide range of errors.

Discussion

The prediction that sophisticated *Ss* will make more precise and less variable errors than naive *Ss* is borne out by the results. The magnitude of the variance of the errors of the sophisticated group permits the inference that members of this group accepted and used the criterion of verticality demonstrated by *E*. The variance of the errors of the naive group does not permit this inference to be drawn, and inspection of the individual results indicates that members of this group failed to maintain any single criterion of verticality, and that some *Ss* changed their criterion even under the same condition.

These results throw light upon the discrepant results reported by Witkin and Asch (6) and by Mann (4) in their series of experiments on the judgment of the vertical. Witkin uses the conflicting visual-proprioceptive tilt chair-room conditions as an unstructured stimulus situation in which the relatively naive *S* is required to choose a vertical. This sets up an ambiguous situation forcing *S* to make his choice from one or more apparent verticals or from the gravitational vertical. Under this condition, the choice is likely to be influenced by attitudes, interests, and by personality differences. The wide range of errors and the significant individual differences reported by Witkin and Asch are probably a function of the level of instructions given *S*.

In the experiments reported by Mann, emphasis during the instructional period is placed upon the gravitational vertical. This is demonstrated by *E*, the *S* practices setting the gravitational vertical, and throughout the experimental series he is frequently reminded to maintain this as his criterion of verticality. The purpose of these experiments is to determine the effect of variation of visual and proprioceptive stimuli upon the judgment of the gravitational vertical, and the instructions are designed to produce and maintain in *S* the direction of the gravitational vertical as the criterion of judgment. Under this more highly structured condition the situational choice is restricted, and the errors reported in Mann's series are smaller and less variable than those reported by Witkin and Asch.

The experimental series reported by these two investigators over the past few years using similar equipment and apparently similar design turn out to be different in purpose, divergent in method, and productive of dissimilar results and conclusions. A comparison of the results of naive and sophisticated *Ss* used in this experiment indicates that the discrepant results reported in these two series of experiments are probably due to differences in levels of sophistication of *Ss* deliberately and justifiably produced by the investigators in their attempts to solve two different problems by the use of similar equipment.

Summary

This experiment was designed to test the effect of instructions upon the perception of the gravitational vertical. In various positions of lateral inclination *Ss* were required to set a target rod to the gravitational vertical against a visual framework.

A naive group of *Ss* was instructed

without amplification to set the target to the vertical. A sophisticated group was required to set the target rod to the gravitational vertical, was practiced in the performance and frequently reminded during the experimental series to maintain the gravitational vertical as the criterion of judgment.

The results indicate that the judgments of the naive group were less precise and more variable than those of the sophisticated group.

It is suggested that the discrepant results reported by Witkin and Asch and by Mann may be due to different levels of sophistication of *Ss* deliberately and justifiably produced by the instructions given by the experimenters in their attempts to solve two different problems.

(Received for early publication
August 28, 1952)

References

1. Boring, R. O. The effect of visual stimulus variables upon the perception of the visual vertical. U. S. Naval Sch. Aviat. Med. and Tulane Univer. Joint Report No. 28, 1952.
2. Gibson, J. J. The relation between visual and postural determinants of the phenomenal vertical. *Psychol. Rev.*, 1952, **59**, 370–375.
3. Gibson, J. J., & Mowrer, O. H. Determinants of the perceived vertical and horizontal. *Psychol. Rev.*, 1938, **45**, 300–323.
4. Mann, C. W. Visual factors in the perception of verticality. *J. exp. Psychol.*, 1952, **44**, 460–464.
5. Wertheimer, M. Experimentelle Studien über das Sehen von Bewegung. *Z. Psychol.*, 1912, **61**, 161–265.
6. Witkin, H. A., & Asch, S. E. Studies in space orientation. IV. Further experiments on perception of the upright with displaced visual fields, *J. exp. Psychol.*, 1948, **38**, 762–782.

Journal of Experimental Psychology
Vol. 45, No. 1, 1953

EFFECTS OF INTENSE PURE TONE STIMULI WHEN MAGNITUDE OF INITIAL INJURY IS CONTROLLED [1]

IRVING E. ALEXANDER AND FREDERICK J. GITHLER

Princeton University

In 1890 Habermann (7) published a study in which, in a single case, high-tone hearing loss was correlated with injury to the basal region of the cochlea. Since that time stimulation-deafness studies have been undertaken with the hope of relating the analysis of pitch to a place principle in the cochlea. Indeed, the German investigators in the early years of this century became embroiled in a controversy regarding the theoretical significance of stimulation-deafness results. Their studies, which generally showed restricted basal end lesions for the high tones and much more widespread injury for the low tones, left much to be desired experimentally. Control of stimulus conditions was generally poor and tests of sensitivity were either inadequate or absent. For a thorough review of the literature up to 1935 the reader is referred to Kemp (8).

In the last 25 years, with improved apparatus and better sensitivity tests, interest in this problem has been renewed. Now in addition to an evaluation of the effect of trauma on cochlear structures one may use the conditioned-response technique, the electrical potentials of the cochlea, or auditory nerve potentials as sensitivity measures of the state of the receptors, the nerve, or both.

Despite the technical improvements there is still not a great deal of agreement in interpretation of stimulation deafness results. Some investigators (**12, 6**) have reported their findings as favorable to a place hypothesis, while others (**13, 3**) have been somewhat less optimistic. These differences sometimes extend beyond the interpretation of theoretical issues and one may find antithetical results reported.

There are two major sources of difficulty in assessing the recent evidence and these perhaps are responsible for the apparent equivocation. In almost all the studies considerable differences in individual susceptibility to trauma were found. Since these studies generally dealt with few *S*s, the wide variability tended to obscure any particular effects of the experimental variable, i.e., loud sound. As a consequence of this, great latitude was afforded in the interpretation of the findings.

Secondly, most investigators have been concerned with structural injury and its correlates with sensitivity in the acute stage, that is, shortly following the exposure. When such measurements are made, one is struck with the apparent lack of correlation between structural damage and sensitivity indices. Indeed, this state of affairs has led some investigators to place little faith in the measurement of cochlear potentials (**9**). The present authors, however, have recently reported closer agreement between changes in electrical sensitivity and

[1] From the Princeton Psychological Laboratory. This research was supported in part by the Office of Naval Research, under Contract N6-onr-270, T. O. 3, Project NR 140-322, and also in part by Higgins funds allotted to Princeton University. Permission is granted for reproduction, translation, publication, use and disposal in whole or in part by or for the United States Government. This research was reported at the 1952 meetings of the Eastern Psychological Association.

TABLE 1

IMMEDIATE LOSS IN SENSITIVITY EXPRESSED IN DECIBELS

Injury Stimulus	S	Test Stimulus (cps)									
		200	300	500	700	1000	2000	3000	5000	7000	10000
300 cps	759	81	66	76	67	75	77	77	75	75	74
	762	85	88	111	112	115	107	95	90	92	70
	766	60	61	70	63	67	77	63	83	90	72
	767	53	66	42	54	46	44	51	69	69	65
5000 cps	768	58	56	64	71	58	60	60	58	56	57
	780	48	48	62	65	58	62	54	58	59	39
	782	58	59	74	65	59	61	62	61	59	59
	783	53	68	69	74	64	61	62	63	60	54

degree of structural damage when time is allowed for the injury process to run its course (**2, 3**).

In the light of the above considerations, the aim of the present study is to introduce two improvements in the usual stimulation-deafness methodology with the hope of clarifying the existing picture. The attempt was made to reduce individual differences by inducing a known decrement in sensitivity. A test-retest procedure was employed to assess both the immediate and more chronic effects of intense pure tones.

PROCEDURE

Guinea pigs were used as *Ss*. Using a sterile technique, a lateral approach to the auditory bulla was made and cochlea potentials were read from the round window membrane. A more detailed report of the procedure may be found elsewhere (**1, 13**). A 1-microvolt response audiogram was determined for the experimental ear. This is a plot of the sound pressures needed to produce a 1-microvolt response at the various frequencies. The ear was then stimulated with one of two test tones, 300 cycles or 5000 cycles at a sound pressure of 1000 dynes/cm.[2]. These tones were chosen because they represent different ends of the audiogram and because comparison data from acute studies were available. The sound was maintained long enough to induce a decrement of 60 db in the sensitivity to the exposure tone. When this decrement had been established, the sensitivity of the ear was again determined for the remainder of the audiogram. The incision was then sutured and the animal was allowed to recover.

When a minimum of three weeks had elapsed, the exposed ear was again tested for sensitivity to tones from 200 cycles to 10,000 cycles. A 1-microvolt response contour was also determined at this time for the nonstimulated ear. This provided a limited check on the general effects of the initial operation on the normal ear.[2]

When all measurements had been made, the animal was sacrificed and the temporal bones removed for later histological evaluation. The entire histological procedure has been previously described in detail (**2, 10**).

In all, eight *Ss* were tested, four in each exposure-tone group, and seven were carried through the histology.

RESULTS

Let us first look at the initial effects of overstimulation with a 300-cycle and a 5000-cycle tone. Table 1 shows the difference in sensitivity between the prestimulation and poststimulation audiograms. For the 300-cycle stimulus injury occurs rapidly. In three of the four cases it was possible to restrict the loss at the exposure frequency to approximately 60 db. The other ear was more seriously injured. The exposure times to produce these decrements ranged from 3 to 6 min. From the individual records it can be seen that the 300-cycle tone

[2] No operative effects were noted. The control ear audiograms were similar to those of the exposed ears before stimulation.

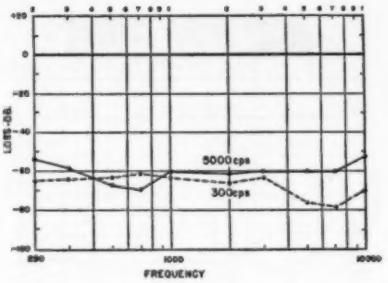

FIG. 1. Composite audiograms determined immediately after exposure to either a 300-cycle or a 5000-cycle tone

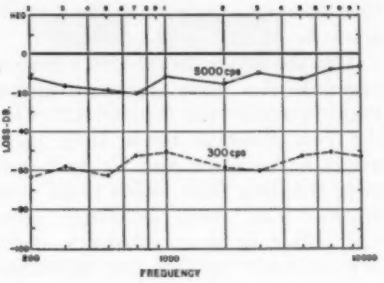

FIG. 2. Composite audiograms determined three weeks after exposure to either a 300-cycle or a 5000-cycle tone

spares no part of the auditory range. All test tones were impaired. Of the three ears in which the criterion (60 db loss) was closely met, two (759 and 765) show similar patterns while the other (767) is in general less severely impaired.

It was a much simpler task to restrict the sensitivity loss to 60 db. for the exposure frequency when 5000 cycles was used as the stimulus, for here injury is more gradual. In every case this restriction was accomplished within 4 db and in a period of time ranging from 18 to 30 min. Again, as for the low-tone stimulus, all tones were affected by this overstimulation. In all cases the range of individual differences was restricted.

The differential effects are presented in Fig. 1, which shows the average loss curve for each of the stimulus tones. In the 300-cycle curve the results from the ear injured far beyond the criterion level are not included. Taking each curve separately we see little variation in the effects of the low-tone stimulus between 200 cycles and 3000 cycles. The curve for this region is almost flat. Sensitivity for the high tones, however, from 5000 cycles to 10,000 cycles, is somewhat more impaired. For the 5000-cycle stimulus the curve remains reasonably flat throughout the range with a slightly greater effect shown at 500 cycles and 700 cycles. The similarity between

TABLE 2

SENSITIVITY LOSS IN DECIBELS THREE WEEKS AFTER STIMULATION

Injury Stimulus	S	Test Stimulus (cps)									
		200	300	500	700	1000	2000	3000	5000	7000	10000
300 cps	759	73	70	79	72	68	80	80	74	72	73
	762	63	64	72	75	72	83	75	78	74	78
	766	73	77	83	70	70	80	81	70	71	80
	767	44	26	23	14	13	13	17	13	8	3
5000 cps	768	17	20	23	25	19	19	22	16	3	7
	780	2	1	7	9	1	2	8	4	1	4
	782	8	1	2	6	10	2	1	3	0	3
	783	40	43	39	41	32	37	24	24	25	21

the two curves for most of the range is marked.

This similarity disappears when we examine the more chronic results. Figure 2 depicts the average curves for the two stimulus tones after three weeks had elapsed and Table 2 contains the data from which these were drawn. The trend is easily seen. The 300-cycle curve shows a slight improvement only in the high-tone range. As a result all the points on the curve fall within a 13 db range. Throughout the range there is still a severe decrement in sensitivity.

The situation has changed radically for the 5000-cycle ears. Improvement has been general and extensive for all tones. While the two curves retain their characteristic flatness, they are no longer dealing with the same magnitudes of loss in sensitivity.

Average curves based on arithmetic means of decibel loss are subject to distortion when there is wide variation in the original data. Let us, therefore, examine Table 2 for another view of the chronic results.

Of the four 300-cycle ears investigated, two (759 and 766) show no recovery at all while the ear which was so extensively damaged initially (762) "recovers" but is still 10 to 20 db poorer than the criterion level. The fourth ear (767) shows recovery in the middle- and high-tone range to within 20 db of the prestimulation levels. This was the ear that was previously described in the poststimulation testing as least impaired in its group.

In the 5000-cycle series, two of the four (780 and 782) recover completely to the prestimulation levels. The other two (768 and 783) have made partial recoveries and are distinctly better in sensitivity throughout the audiogram than they were immediately following stimulation. It is apparent, then, for a constant loss of 60

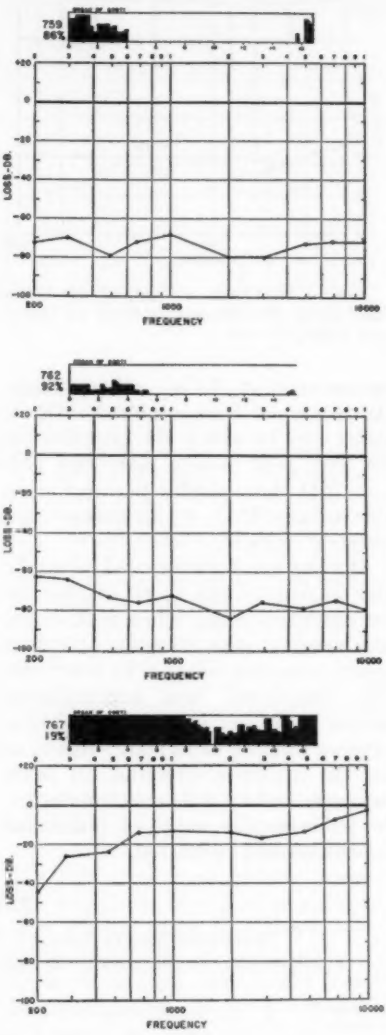

FIG. 3. Linear transformations of the state of the organ of Corti and the corresponding electrical audiograms for ears of the 300-cycle group. These measurements reflect the condition of the ears three weeks after the exposure. The animal number and percentage of damage accompany the organ of Corti map.

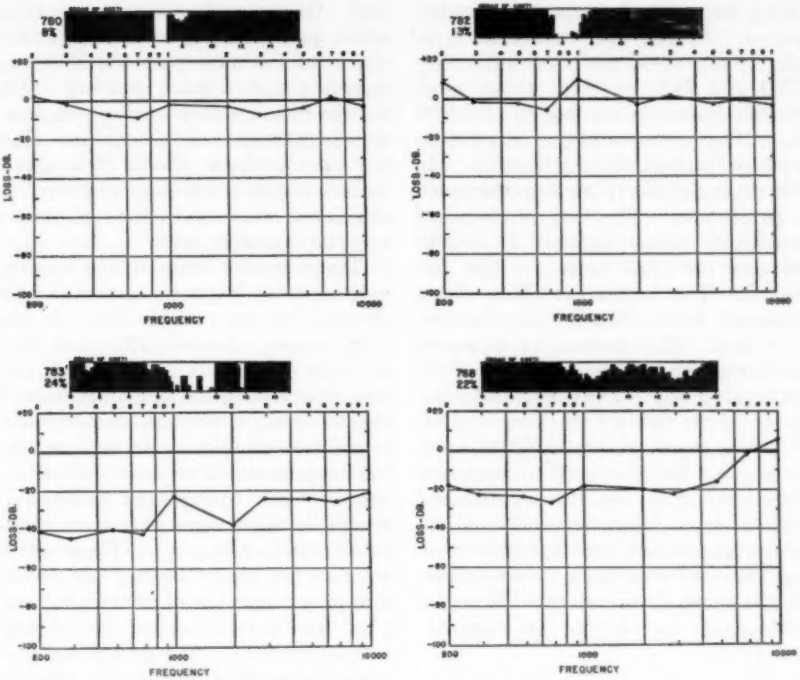

FIG. 4. Linear transformations of the state of the organ of Corti and the corresponding electrical audiograms for ears of the 5000-cycle group. These measurements reflect the condition of the ears three weeks after the exposure. The animal number and percentage of damage accompany the organ of Corti map.

db, one may expect recovery when 5000 cycles is the stimulating tone while the probability of recovery is diminished seriously when 300 cycles serves as the stimulus.

We can now turn to the histological findings. Figure 3 contains the effects of a 300-cycle injury. In each case is presented a linear transformation of the organ of Corti and an evaluation of its structures.[3] In addition the electrical audiogram at the time the animal was sacrificed is presented for comparison. Of the three cases that were carried through the histology, two (759 and 762) show almost complete destruction of the organ of Corti structures except in the first few millimeters of the basal turn. The electrical audiograms are also most severely depressed. The third (767) shows moderate to severe injury starting at 8 mm. and continuing almost to the apex. The corresponding electrical audiogram shows a moderate decrease in sensitivity for most of the range and almost complete recovery at 7000 cycles and 10,000 cycles.

The histological results for the 5000-cycle stimulus are presented in Fig. 4

[3] A value of 20% was assigned to each row of hair cells, four in all, and the remaining 20% was distributed among the other organ of Corti structures. Damage to the auditory nerve fibers and ganglion cells was also assessed but is not reported here.

along with the corresponding audiograms. In the two ears that completely recovered electrical sensitivity (780 and 782) we find rather small specific lesions in the region of 6 to 8 mm. Here are specific cochlear lesions with normal cochlear potentials. In 783 the main effects are found between 6 to 10 mm. The loss in electrical sensitivity ranges between 21 and 43 db and the high tones are less impaired. The fourth ear (768) shows moderate injury from 5.4 mm. through 15.4 mm. The pattern of injury is similar to the one found in 767 of the 300-cycle group. The electrical audiograms of the two are also very similar.

As in our previous work (**2, 3**) when an average figure is used to represent the electrical loss and this is correlated with a single figure representing the percentage of cochlear structures missing, the correspondence is very close. To the seven cases presented here the rank-order correlation is approximately .9.

Discussion

By inducing a sensitivity loss of known magnitude at the exposure frequency we have succeeded in reducing the effect of individual variations in susceptibility to trauma. As a result, the effects of the stimulating tones may be more accurately gauged. For the efficacy of our procedure let us compare our work with previous results. Wever and Smith (**13**) also used a 300-cycle stimulus at a sound pressure of 1000 dynes/cm.^{2} for a constant interval of 4 min. This time interval falls within our range of 3 to 6 min. of stimulation. Yet their acute results range in variability from no apparent effects to ones that reach and exceed our criterion. One might argue that our acute results at 300 cycles appear uniform because of the magnitude of the sensitivity loss imposed on the ear. On the surface there seems to be some justification for this argument since for the most part the 300-cycle ears do not show much recovery. The results from the 5000-cycle cases, however, run counter to this notion. Here we have uniform effects of a severe nature in the acute stage and varying degrees of recovery, including a return to prestimulation levels.

These results support the contention that the low tones inflict greater damage to the cochlea than do the high tones. Smith (**10**) took this position on the basis of his work and thus postulated that the amplitude of displacement of the basilar membrane is the critical variable in determining the magnitude of structural damage. Davis *et al.* (**6**) reported findings in which the high tones were more effective in causing damage. These investigators felt that this was due to the damping properties of the tensor tympani muscle as it carried out its protective function. We find this explanation somewhat tenuous in view of the fact that overstimulation was carried on for protracted periods ranging from 16 to 24 hr. a day for many days.

With regard to auditory theories we find in this work, as in our noise studies, the weight of the evidence against a place hypothesis. In our two fully recovered ears (780 and 782) we find upper basal lesions which cover approximately 10% of the cochlea and no corresponding diminution in cochlear potentials. Were we to be guided by the various cochlear localization maps that have been proposed (**4, 5, 11**), we should expect to find marked impairments of sensitivity in the region of 3000 cycles. That this is not the case is obvious from the sensitivity curves presented.

In addition, we find similar histological and electrical injury when tones varying widely in frequency are

used as producers of that injury (767 and 768). Certainly, if the cochlea is selectively tuned in other than a very gross fashion, we should not expect such a result. A shred of evidence may be gathered for high-tone localization from these same ears where there is little loss at the high tones and no injury for most of the basal turn. But even such evidence leaves one in a quandary when in the same study one finds cases in which basal structures are left intact but sensitivity loss extends throughout the range (759, 762, 783).

One further fact emerges from this study which we are at a loss to explain but which is also not commensurate with place theory expectations. Acute sensitivity impairment to a low-tone stimulus is greatest for the extreme high-tone range; i.e., 5000–10,000 cycles. This same finding was reported earlier by Wever and Smith (**13**).

Summary

Guinea pigs were subjected to overstimulation from either a 300-cycle or a 5000-cycle tone for a period long enough to produce a 60 db loss in sensitivity for the stimulating tone. Cochlear potentials were taken prior to stimulation, immediately following, and after a lapse of three weeks. The ears were also studied histologically. The findings are as follows:

1. Individual variability in susceptibility to trauma is reduced under these conditions of stimulation.

2. Test-retest measurement of cochlear potentials in the guinea pig is feasible.

3. The acute results show that for a 300-cycle stimulus all tones up to 5000 cycles are affected to the same degree. The highest tones tested are more seriously impaired. For a 5000-cycle tone the entire audiogram suffers an approximately equal loss.

4. After three weeks some recovery of function takes place in all of the 5000-cycle ears. Two of the group returned to normal levels.

5. The 300-cycle injury is much more permanent. It appears that at equal sound pressures a low-tone stimulus is much more effective in producing injury than a high-tone stimulus.

6. Cochlear injury for a 5000-cycle stimulus is generally more restricted than for a stimulus of 300 cycles.

7. For the cases reported here there exists a rank-order correlation of approximately .9 between the average loss in the electrical audiogram and the percentage of cochlear structures destroyed.

The results were discussed with reference to a place principle in the cochlea and were interpreted to be generally unfavorable to such a principle.

(Received July 26, 1952)

References

1. Alexander, I. E., & Githler, F. J. The effects of jet engine noise on the cochlear response of the guinea pig. *J. comp. physiol. Psychol.*, 1949, **42**, 517–525.
2. Alexander, I. E., & Githler, F. J. Histological examination of cochlear structure following exposure to jet engine noise. *J. comp. physiol. Psychol.*, 1951, **44**, 513–524.
3. Alexander, I. E., & Githler, F. J. Chronic effects of jet engine noise on the structure and function of the cochlear apparatus. *J. comp. physiol. Psychol.*, 1952, **45**, 381–391.
4. Culler, E. A. An experimental study of tonal localization in the cochlea of the guinea pig. *Ann. Otol., etc., St. Louis*, 1935, **44**, 807–813.
5. Culler, E. A., Coakley, J. D., Lowy, K., & Gross, N. A revised frequency-map of the guinea pig cochlea. *Amer. J. Psychol.*, 1943, **56**, 475–500.
6. Davis, H., Derbyshire, A. J., Kemp, E. H., Lurie, M. H., & Upton, M. Functional and histological changes in the guinea pig resulting from prolonged stimulation. *J. gen. Psychol.*, 1935, **12**, 251–278.

7. Habermann, J. Ueber die Schwerhörigkeit der Kesselschmiede. *Arch. Ohrenhk.*, 1890, **30**, 1–25.
8. Kemp, E. H. A critical review of experiments on the problem of stimulation deafness. *Psychol. Bull.*, 1935, **32**, 325–342.
9. Office of Scientific Research and Development. Committee on Medical Research Report. Supplementary report on injury of the inner ear produced by exposure to loud tones. December 31, 1943.
10. Smith, K. R. The problem of stimulation deafness. II. Histological changes in the cochlea as a function of tonal frequency. *J. exp. Psychol.*, 1947, **37**, 304–317.
11. Stevens, S. S., Davis, H., & Lurie, M. H. The localization of pitch perception on the basilar membrane. *J. gen. Psychol.*, 1935, **13**, 297–315.
12. Upton, M. Functional disturbances of hearing in guinea pigs after long exposure to an intense tone. *J. gen. Psychol.*, 1929, **2**, 397–412.
13. Wever, E. G., & Smith, K. R. The problem of stimulation deafness. I. Cochlear impairment as a function of tonal frequency. *J. exp. Psychol.*, 1944, **34**, 239–245.

Journal of Experimental Psychology
Vol. 45, No. 1, 1953

LEARNING AND PERFORMANCE IN EYELID CONDITIONING AS A FUNCTION OF INTENSITY OF THE UCS [1]

KENNETH W. SPENCE

State University of Iowa

Experimental studies concerned with the effects upon behavior of variation of the motivational level of S may be grouped into three classes. The first class contains those experiments in which after training in the situation under some constant period of deprivation (T_D) a test or measurement period is given in which T_D is varied. These studies are regarded as providing a measure of the drive factor (D) that determines response strength and the empirical relation $R = f(T_D)$ is taken as a reflection of the relation of D to T_D. The classical obstruction box studies of Warden and his associates (**19**) and more recent investigations (**4, 8, 12, 13, 20**) in which the Skinner box type of situation was used fall into this class.

In the second class of experiments T_D is varied (i.e., is different for different subgroups) during the first training period and then a second test period is given in which T_D is the same for all groups. The objective of these studies has been to ascertain whether the learning factor (habit strength) is a function of the differences in drive strength resulting from different values of T_D during the first period. Instances of experiments designed in this fashion are those of Finan (**3**), Kendler (**7**), and Strassburger (**15**). All of these studies used simple instrumental conditioning.

A third group of studies has not attempted to differentiate between the two factors of habit (H) and drive (D) but has simply studied performance during the training period as a function of differences in T_D. Typically these investigations have involved more complex learning situations such as the maze (**9, 16, 17, 18**) or discrimination learning (**2, 10**). Recently, however, two studies of this type have been reported in which classical defense conditioning was employed (**11, 14**). In these studies the experimentally manipulated motivational variable was the intensity of the unconditioned stimulus (UCS) rather than the time of deprivation of some goal object.

This third group of studies, it should be noted, does not permit one to identify which factor, learning (habit strength) or drive strength, lies behind the obtained differences in performance level. Quite frequently, however, these investigators have inferred that the acquisition process (habit formation) was superior under conditions of strong motivation. Thus, Passey (**11**) interpreted his finding that the level of conditioning was a positive function of the intensity of the UCS as supporting Hull's assumption (**5**) that more habit strength was developed with a stronger drive. He pointed out that this was in line with Hull's special reinforcement hypothesis that related the increment of habit strength per trial to the amount of drive reduction or drive stimulus reduction.

While it is true that this finding is in agreement with the implication of

[1] The writer wishes to acknowledge the assistance of George Harker, Henry Loess, Wallace McAllister, Donald Swisher, and Janet Taylor who served as assistants either during preliminary experimentation or the present study.

Hull's hypothesis, it should be noted that there are two other possible interpretations of it even within the framework of Hull's theorizing. Thus, according to this formulation, response strength is a function of both habit and drive strength, i.e., $R = f(H \times D)$. It is conceivable that the increments of habit strength (H) do not vary with the intensity of the UCS and T_D but that only drive strength (D) does. Finally, there is the possibility that differences in the intensity of the UCS or in T_D might result in differences in both of these hypothetical factors, H and D. The experimental design employed by Passey does not permit one to make any decision among these alternatives.

The appropriate experimental design with which to investigate this problem is the factorial type shown later in Table 1. The experiment is conducted in two periods. In the first the *S*s are divided randomly into two groups, one of which is conditioned with a weak UCS and the other with a strong UCS. In the second period these two groups are each divided into two subgroups with half of the *S*s continuing with the same UCS as in the first period and half being switched to the other intensity. There are thus four different subgroups in this second experimental period.

Response measures obtained in the second period provide the data. Comparison of the row means indicates whether response strength during this period varies with the level of intensity of the UCS present in the first period. As the level of intensity of UCS in the second measuring period is controlled, this comparison cannot reflect differences in drive strength (D) resulting from a difference in intensity of UCS. Hence, a significant difference in these row means would be interpreted as indicating that a different amount of habit strength developed in the two groups during the first period. An accompanying significant difference in the column means would be interpreted as reflecting differences in drive strength (D) resulting from the different intensities of UCS present at the time of measurement. Presumably such a difference would also reflect differences in habit strength that would develop during the second period. If, on the other hand, there is no difference in the row means and a significant difference in the column means, this would lead to the interpretation that habit growth was not a function of the motivational level of the organism and that the difference between the groups under different intensities of UCS reflected differences in drive strength (D) only. The absence of a significant difference in the column means and a significant difference in the row means would imply that the difference in response strength reflected different habit strengths with no difference in drive strength.

Attention should be called to the fact that the studies in the first two of our three classes of motivational experiments conform partly in design to the factorial type described above except that they have been incomplete designs. Thus, the investigations in the first group had but a single motivational condition in the first training period and two or more different motivational conditions in the second period. In terms of Table 1 there was but a single row and a number of columns consisting of but a single cell. Studies in the second group, on the other hand, have typically had but a single column with a number of different rows each with but a single cell. Only the Kendler study involved the complete factorial design.

The partial nature of the experimental design employed in these investigations has resulted in a certain amount of distortion in their findings. Thus, in the first class of studies, those concerned with differences in drive strength (D), the typical procedure has been to carry on the original training in the setup with a deprivation interval of approximately 24 hr. Subsequent comparisons of response strength at different deprivation periods, e.g., 3, 6, 12, and 24 hr., do not make proper allowance for the fact that the 3-, 6-, and 12-hr. groups are performing at *different* conditions from those employed in the first period while the 24-hr. group has the *same* deprivation condition. Interpreted in terms of our theoretical constructs H and D, this design provided assurance that a constant amount of H was *developed* in the groups to be run subsequently under the different deprivation intervals. It did not, however, control for the possibility that a significant part of the stimulus complex that acquires habit strength for the response consists in the internal cues from the need state (S_D). If these cues are discriminably different for the different deprivation periods, and there is experimental evidence (**1, 6**) to indicate that they are, then the stimulus complex for any groups run under a different deprivation condition in the test period from that used in the original training period is changed. This would mean that the effective habit strengths available for the different groups during the test periods would not be constant but would differ from each other in varying amounts that would be a function of the amount of change in the stimulus complex (principle of stimulus generalization). *So far as*

providing the precise nature of the function relating D *to* T_D, these studies have been worthless because they have not controlled for these differences in generalized habit strength.[2]

The present experiment attempted to investigate the question as to whether differences in response strength in classical defense conditioning resulting from the use of different intensities of the UCS are to be interpreted as reflecting, in part at least, differences in the strength of conditioning, i.e., habit strength (H), or whether the differences reflect only differences in drive strength (D). The experimental design employed was the factorial type discussed above.

Procedure

Subjects.—A total of 97 men from an introductory course in psychology served as *Ss*. Women were not used as it was found in an exploratory experiment that a very high proportion of them responded with voluntary blinks to the CS under the condition in which the UCS (air puff) was very strong (5.0 lb./sq. in.). Eleven men were eliminated because of the fact that they gave a very high incidence of voluntary closures to the conditioned light signal. Such voluntary responses were distinctly different from conditioned responses both in their form and in their latency distribution. Thus, the voluntary pattern of response, consisting of a smooth, sharp closure which was maintained until after the onset of the air puff, occurred almost exclusively in the period 200 to 460 msec. following the onset of the CS. The responses predominating in the period 460 to 755 msec. were more gradual and much more irregular in their pattern. All *Ss* who had a high incidence (one-third or more) of responses with latencies less than 460 msec. in either one of the two conditioning periods were eliminated from the experiment. Questioning of these *Ss* revealed that there was not always awareness that such "voluntary" closures were taking place. Six *Ss* were eliminated from the experiment because of error on the part of *E* or apparatus difficulties. Eighty *Ss* remained for the experiment described below.

Apparatus and method of recording.—The *S* was seated in a dental chair in a semidarkened room adjoining that in which the recording apparatus and stimulus controls were located. Fixed in a reclining position in the chair, *S* was instructed to blink on receiving a ready signal, and then to fixate on a 6-in. circular milk glass disc, placed at a distance of 132 cm. The brightness of the disc between trials was .05 apparent ft.-candles.

The CS consisted in an increase in brightness of the disc to 1.51 apparent ft.-candles. The duration of the CS on each trial was 825 msec. The onset and duration of the CS and UCS were controlled by means of Hunter electronic timers and were recorded on a Brush polygraph by means of a Brush BL-902 penmotor.

The UCS, a puff of air, followed the onset of the CS by 755 msec. and was delivered to the right eye. The source of the air puff was a high pressure compressed air line provided in the building. Appropriate pressure was obtained by means of a variable (0–10 lb./sq. in.) air pressure reducing valve. A pressure gauge was used to provide the two strengths of air pressure employed, .25 lb./sq. in. in the case of the weak UCS and 5.0 lb./sq. in. in the case of the strong UCS. A 110-v., 60-cycle AC operated solenoid valve, controlled by an electronic timer, limited the duration of the puff to 50 msec. This represents a much superior air-puff system than that previously employed in this laboratory in that the air pressure can be maintained at the stated pressure for any desired time interval.

The movement of the eyelid was recorded by a system similar to that described earlier by Spence and Taylor (**14**) except that a simpler method of amplification was employed. Briefly the technique involved the use of (*a*) a lightweight microtorque potentiometer, (*b*) a direct current amplifier, (*c*) a Brush polygraph, and (*d*) a Brush BL-902 penmotor.

The microtorque potentiometer was mounted on an adjustable headband just above *S*'s eye. A very light lever was attached at right angles to the variable (rotating) arm of the potentiometer and a mechanical linkage was accomplished between it and a tiny plastic lever fastened to the upper eyelid by adhesive tape. A small restoring spring was added to the potentiometer arm thus permitting it to follow the eyelid's motion. The system was very light and involved a minimum of work. Movement of the eyelid, with its accompanying rotation of the potentiometer arm, altered the DC voltage appearing on the grids of a push-pull vacuum tube amplifier. The latter was a three stage push-pull amplifier with a push-pull cathode follower output stage

[2] Yamaguchi's study is an exception to this in that in it the time of deprivation employed in the original training period was the same as that used in the test period. In effect Yamaguchi took advantage of the Strassburger and Kendler findings that T_D did not effect the amount of H developed in the first period.

being used to match the impedance of the Brush penmotor.

Conditioning procedure.—The experiment was conducted in two periods, the second following the first by 24 hr. Following the reading of instructions each *S* received three presentations of the light alone in order to check for any initial conditioned response tendency to the light. A single air puff was then administered and the eyelid behavior recorded for 40 sec. All *S*s were then given 30 paired presentations of the light and air puff. Intertrial intervals of 15, 20, or 25 sec., averaging 20 sec. and arranged according to a fixed schedule, were used.

On this first day half of the 80 *S*s were conditioned with the relatively weak air puff (.25 lb./sq. in.) and half with the strong puff (5.0 lb./sq. in.). At the beginning of Day 2 the instructions were repeated and a further 70 paired presentations of the CS and UCS were given. On this day half of the *S*s trained with the weak puff on Day 1 were trained for 70 more trials with it and half were shifted to the strong UCS. Similarly, half of the *S*s trained on Day 1 with the strong UCS had the strong UCS continued on Day 2 and half were shifted to the weak UCS. Thus, on Day 2 there were four subgroups of 20 *S*s each in terms of the UCS used in the two conditioning periods: WW, WS, SS, and SW.

At the end of the session all *S*s were asked not to discuss the experiment with other members of the class.

Results and Discussion

The values in the four cells of Table 1 represent the mean number of CR's made by the four subgroups in the first 20 trials of the second day. This 20-trial period was chosen because a longer period ran into the problem that our measuring scale has a ceiling, i.e., 100%. The frequency measure of response strength is quite satisfactory only so long as we do not reach this ceiling level. Once an *S* has reached 100% any further increase in strength of the *S–R* tendency ($_sE_R$) with further training cannot be reflected by this measure. By the second ten-trial block two *S*s were responding 100%, and in the next ten-trial block four more *S*s reached this level. Accordingly, it was decided to use the data for the first 20 trials only. The scores obtained in this period provided quite satisfactory distributions so far as the requirements (variance and form) for statistical analysis are concerned. Score distributions for longer periods were quite unsatisfactory, for with the ceiling on the measure the distributions became more and more skewed as more *S*s attained the 100% level of response.

Returning to Table 1, it will be seen that there are sizable differences both in the case of the row means and the column means. The statistical significance of these variations was tested by means of analysis of variance, the results of which are presented in Table 2. It will be noted that there was no significant interaction. The "main effect" of the intensity of the UCS on Day 1, as shown by the difference between the two row means, is revealed to be significant by the *F* test at a level just short of the .01 level of confidence. The difference between the column means, which reflects the "main effect" of the intensity of UCS at the time of measurement (Day 2), is also seen to be highly significant by the *F* test, in this instance beyond the .001 level of confidence.

The significant difference between the row means supports the interpretation, it will be recalled, that different strengths of conditioning (amounts of

TABLE 1

Mean Number of CR's Made in First 20 Trials of Day 2

Day 1 UCS lb./sq. in.	Day 2 UCS (lb./sq. in.)		Means Reflecting *H*
	.25	5.00	
.25	5.65	8.80	7.23
5.00	7.45	13.00	10.23
Means Reflecting *D*	6.55	10.90	

TABLE 2

ANALYSIS OF VARIANCE FOR TRIALS 31–50 (DAY 2)

Source	*df*	Mean Square	*F**
Intensity Day 1 (rows)	1	180.00	6.40
Intensity Day 2 (columns)	1	378.45	13.46
Rows × columns (interaction)	1	28.85	1.03
Within groups	76	28.11	
Total	79		

* .05 level, $F = 3.98$; .01 level, $F = 7.01$.

habit strength) were developed during the first day of training when the UCS was different for the *S*s compared. This result would be in line with the notion that the increment of habit strength (ΔH) is a function of the strength of drive and interpretable as in agreement with the drive-reduction theory of reinforcement; that is, the stronger UCS would produce a stronger drive and with its cessation there would be a greater drive reduction.

The significant difference between the column means indicates that response strength varied with the intensity of the UCS employed at the time of measurement. Presumably this difference reflects the differences in drive strength (D) produced by the different intensities of UCS. However, if habit strength also varies with intensity of UCS, then there must also be a difference between the habit strengths of the two groups (.25 and 5.0). This difference in habit strength would be that which developed during this 20-trial period of Day 2 for they were equated in terms of the amount of habit strength developed on Day 1. While no evaluation can be made of the differences, it is of some interest that the difference between the column means (4.4) is 47% larger than the difference between the row means (3.0). The former involves only a difference in H based on 30 trials with different intensities of UCS.

A problem that arises in connection with this interpretation of a different amount of habit strength being built up with the different intensities of UCS is the negative evidence that has been obtained in studies employing an appetitional need such as hunger. Thus, the experiments of Kendler (**7**) and Strassburger (**15**), which involved varying the time of deprivation of some goal object (food, water), did not provide comparable significant differences in the row means of the response measures in the second period. An answer that suggests itself is that the amount of reinforcement (drive reduction) in the food reward situation is the same for different conditions of deprivation. Indeed, the reinforcement probably is secondary in nature and does not involve any primary drive reduction whatever. In the conditioning study, on the other hand, the reinforcement was primary (pain reduction) and, presumably, different for the different intensities of air puff.

There are, of course, other possible interpretations as to the factor responsible for the difference in the row means, i.e., the difference related to the intensity of the UCS on Day 1. One most likely to be offered by the theorist opposed to any kind of reinforcement notion is that a *bigger* response is probably conditioned in the case of the stronger UCS. The logic is not too clear as to how one gets from this to the predicted greater *frequency* of CR for the group which is making the larger unconditioned response (UCR) although it is evident that a higher average magnitude of CR would be expected. Presumably it is assumed that the magnitude of the CR increases with successive trials and there is thus a greater likelihood of a response reaching the criterion magnitude (1 mm.) of a CR earlier in the case of *S*s with a large UCR than those with a smaller one. An examination of the magnitude of the UCR during the first ten trials of Day 1 did reveal that the mean

UCR was larger for the 5.0-lb. group than for the .25-lb. group (18.17 mm. vs. 14.5 mm.). However, an examination of the scatter diagrams relating amplitude of UCR on these trials with the number of CR's on Day 2 (Trials 31–50) showed no evidence of a positive correlation. In fact the relationship was slightly but insignificantly negative.

As a further means of checking on the possibility that amplitude of UCR was a factor, *Ss* in the four subgroups were matched by selecting ten *Ss* in each subgroup whose mean amplitude of UCR on Day 1 fell between 14 and 20 mm. The mean amplitudes of the four subgroups when so matched were either 17.3 or 17.4 mm. The mean number of CR's in Trials 31–50 on Day 2 for the WW and WS subgroups and for the SW and SS subgroups were then calculated for these *Ss*. They were 6.5 for the *Ss* (WW and WS) who had the weak UCS on Day 1 and 11.5 for the *Ss* (SW, SS) who had the strong UCS. The difference, 5.0, is even larger, it will be noticed, than that which appears between the row means in Table 1.

A second possible interpretation of this finding that would not involve the reinforcement principle is that a stronger emotional or fear reaction was established to the cues from the experimental situation in the case of the more intense UCS. As the conditioned fear response is one of the determiners of the drive level (*D*), it would follow that *Ss* trained with the more intense UCS would be expected to have a higher *D* level. According to this formulation, however, the different levels of response are due to this different level of *D* (from the secondary motivation of fear) and not to a difference in *H*. The present experiment does not, of course, permit an evaluation of these rival interpretations. Its chief contribution is the experimental finding that performance level in the second period is a function of the differential conditions in the first period. This fact points to the necessity for invoking some kind of historical or learning factor, different habit strength, conditional fear responses, or some such to account for it.

Summary and Conclusions

A factorial type experiment was conducted to investigate the question as to whether a difference in the intensity of the UCS (air puff) during a first training period would lead to differences in the level of eyelid conditioning in a second period in which intensity of the UCS was controlled.

During the first experimental period of 30 trials 80 *Ss*, divided randomly into two equal groups, were conditioned. One group was conditioned with a weak UCS and the other with a strong UCS. In a second training period given on the following day the two groups were each divided into two subgroups with half of the *Ss* continuing with the same UCS and half being switched to the other intensity.

Analysis of variance of the frequency of CR's in the first 20 trials of the second period revealed that there were significant differences in performance in relation both to the intensities of UCS employed in the first period and also in relation to that used in the second period. The fact that response level varied with the intensity of the UCS employed on Day 1 was interpreted as in accord with Hull's theory that the increment of habit strength is a function of the amount of primary drive reduction. The finding that the level of CR varied with the intensity of the UCS at the time of measurement on Day 2 was taken as a reflection of difference in drive level (*D*). Other possible interpretations not involving the principle of reinforcement were also suggested.

(Received July 21, 1952)

References

1. Bloomberg, R., & Webb, W. B. Various degrees within a single drive as cues for spatial response learning in the white rat. *J. exp. Psychol.*, 1949, **39**, 628–636.
2. Dodson, J. D. Relative values of reward and punishment as habit formation. *Psychobiol.*, 1918, **1**, 231–276.
3. Finan, J. L., & Taylor, L. F. Quantitative studies in motivation. I. Strength of conditioning in rats under varying degrees of hunger. *J. comp. Psychol.*, 1940, **29**, 119–134.
4. Horenstein, B. R. Performance of conditional responses as a function of strength of hunger drive. *J. comp. physiol. Psychol.*, 1951, **44**, 210–224.

5. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
6. Jenkins, J. J., & Hanratty, J. A. Drive intensity discrimination in the albino rat. *J. comp. physiol. Psychol.*, 1949, **42**, 228–232.
7. Kendler, H. H. Drive interaction: II. Experimental analysis of the role of drive in learning theory. *J. exp. Psychol.*, 1945, **35**, 188–198.
8. Kimble, G. A. Behavior strength as a function of the intensity of the hunger drive. *J. exp. Psychol.*, 1951, **41**, 341–348.
9. MacDuff, M. M. The effect on retention of varying degrees of motivation during learning in rats. *J. comp. Psychol.*, 1946, **39**, 207–240.
10. Meyer, D. R. Food deprivation and discrimination reversal by monkeys. *J. exp. Psychol.*, 1951, **41**, 10–16.
11. Passey, G. E. The influence of intensity of unconditional stimulus upon acquisition of a conditional response. *J. exp. Psychol.*, 1948, **38**, 420–428.
12. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–109.
13. Sackett, R. S. The effect of strength of drive at the time of extinction upon resistance to extinction in rats. *J. comp. Psychol.*, 1939, **27**, 411–431.
14. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
15. Strassburger, R. C. Resistance to extinction of a conditional operant as related to drive level at reinforcement. *J. exp. Psychol.*, 1950, **40**, 473–487.
16. Teel, K., & Webb, W. B. Response evocation on satiated trials in the T-maze. *J. exp. Psychol.*, 1951, **41**, 148–152.
17. Tolman, E. C., & Gleitman, H. Studies in spatial learning. VII. Place and response learning under different degrees of motivation. *J. exp. Psychol.*, 1949, **39**, 653–659.
18. Tolman, E. C., & Honzik, C. H. Degrees of hunger, reward and non-reward, and maze learning in rats. *Univer. Calif. Publ. Psychol.*, 1930, **4**, 241–256.
19. Warden, C. J. *Animal motivation: experimental studies on the albino rat.* New York: Columbia Univer. Press, 1931.
20. Yamaguchi, H. G. Drive (*D*) as a function of hours of hunger (*h*). *J. exp. Psychol.*, 1951, **42**, 108–117.

Journal of Experimental Psychology
Vol. 45, No. 1, 1953

PERCEPTION OF THE STATISTICAL STRUCTURE OF A RANDOM SERIES OF BINARY SYMBOLS [1]

HAROLD W. HAKE AND RAY HYMAN

The Johns Hopkins University

Our problem was to learn how human *Ss* perceive the statistical structuring imposed upon apparently random series of binary symbols. The series used were similar to those previously used in studying the effects of partial reinforcement in human verbal learning in that they were chance series constructed by following probability rules determining the appearance of either symbol on each trial (**2, 3, 6**).

It has been shown that *Ss* can learn the probability of occurrence of one of two possible symbols in such a chance series and will predict that symbol about as frequently as it is likely to occur in the series (**2**). These results seem to us to be convincing evidence that *Ss* must be responding to whole series of trials rather than to a single trial as an isolated unit. The stimulus for responding on each trial must be a composite of whatever experiences they have had on previous trials. This is certainly the case in human verbal learning of the type originated by Humphreys where the stimulus on each trial is the series of trials preceding it (**4**). The stimulus on a given trial differs from the stimulus on any other trial only in its temporal position in the trial series.

We chose this experimental situation, therefore, as a good one in which to learn how *Ss* perceptually group together the events occurring in a series of trials as the stimulus for action on a succeeding trial. By introducing sequential dependencies into some of our trial series, we varied the likelihood, from series to series, of a given trial being preceded by a particular combination of symbols. We used four trial series which differed from each other in the probability of occurrence of a particular stimulus on any given trial (first-order probability), and differed from each other also in the degree to which the appearance of one of two possible symbols on a particular trial depended upon which symbol had occurred on the preceding trial (second-order probability). The degree to which the predictions of *Ss* were determined by such first- and second-order statistical structuring in the trial series was analyzed by a method derived from communication theory (**8**).

Method

Apparatus.—The symbol which *Ss* saw on each trial was either a vertical column or a horizontal row of four small lighted neon bulbs. The sequences in which these two symbols appeared were punched in code on teletype tape. This tape was read by an electronic sequence programmer which presented all of the trials automatically. Each trial consisted of a warning light followed 4.5 sec. later by one of the two symbols. The intertrial interval was held constant at 9 sec., and the stimuli remained on for 1.7 sec.

Procedure.—All *Ss* saw a series of 240 symbols presented one after the other. Immediately following the warning light, and before the symbol appeared, on each trial, *Ss* were required to predict which of the two possible symbols would occur on that trial. They made this prediction by checking either *H* or *V* on answer sheets,

[1] This work was supported under Contract N5-ori-166, Task Order I, between the Systems Coordination Division, Naval Research Laboratory, Office of Naval Research, and The Johns Hopkins University. This is Report No. 166-I-160, Project Designation No. NR 507-470, under that contract.

TABLE 1

THE EXPERIMENTAL SERIES

Probability Rules	Series 1	Series 2	Series 3	Series 4
$P(H)$	.50	.50	.75	.75
$P_H(H)$	.50	.80	.75	.90
$P(V)$	.50	.50	.25	.25
$P_V(V)$	.50	.80	.25	.70

They were given no information about the rules by which the series were constructed.

Four groups of *S*s were used, each of which predicted just one series. The four series are summarized in Table 1. In Series 1 the *H* symbol occurred on one-half of the trials. This was also true of Series 2 which had the additional statistical restriction that whenever an *H* had occurred, the probability that it would be followed by another *H* was .80. Series 3 and 4 had *H* symbols occurring on three-quarters of the trials. Series 4 had the additional sequential probabilities shown in the table. All series were constructed following these probability rules which applied to the *V* symbol also as shown in the table. They were constructed so that the rules held in each successive group of 24 trials.

Subjects.—A total of 40 *S*s was used, ten in each group. They were men and women undergraduate students and were run in groups of two or three at a time.

RESULTS

The data of this study are the series of 240 predictions made by each *S*. In each series, every prediction except the first was made following the occurrence of some set of events on previous trials. We have analyzed these predictions to show the degree to which each was determined by these previous events as a function of the number of previous trials being considered and the number of different kinds of events which were considered to have occurred on those trials.

Description of the uncertainty measure.—Our basic measurement is the *uncertainty* measure which has been described fully elsewhere (**8, 9, 10, 11**). The computations of the uncertainty or *entropy* associated with a series of predictions of an *S* came from:

$$H_i(j) = -\sum_i \sum_j p(i) p_i(j) \log_2 p_i(j), \quad (1)$$

where $H_i(j)$ is our uncertainty in stating which symbol *S* will predict on a given trial when we know the previous events described by (i); $p_i(j)$ is the proportion of trials on which *S* predicted a given symbol when he had experienced the previous events (i); and $p(i)$ is the proportion of trials before each of which he had experienced events (i). Since the predictions of our *S*s always involved the choice of one symbol from a set of two, maximum uncertainty is unity and minimum is zero.

This measure has the useful property that

$$H_i(j) \geqslant H_I(j), \quad (2)$$

where $H_I(j)$ is an uncertainty measure based upon the same series of predictions as is equation (1) but where the previous events (i) are more completely known. Expression (2) states that our uncertainty $H_I(j)$, which results from knowing *more* about what *S* has experienced at the time he makes his choice, is less than or equal to our uncertainty $H_i(j)$, which results from knowing *less* about what he has experienced at the time he makes his choice. The two uncertainties can be equal only in the case where increased knowledge of previous events does *not* decrease our uncertainty. When increased knowledge of previous events which *S* has experienced does decrease our uncertainty in stating which symbol he will predict on the next trial, then we assume that these events were discriminated by *S* and did act to determine his choice. We have used this relationship to determine the amount by which our uncertainty, in stating which symbol an *S* will predict on a given trial, is reduced as we define

TABLE 2

THE AVERAGE UNCERTAINTIES RESULTING WHEN THE JOINT EVENTS OF PREVIOUS TRIALS ARE KNOWN

Average Uncertainties	Series 1	Series 2	Series 3	Series 4
$H(y)$	1.00	1.00	.76	.80
$H_{xy}(y)$	.98	.54	.68	.55
$H_{xy,\ xy}(y)$	.95	.52	.66	.54

more and more explicitly the events occurring previous to his prediction. We have expressed this amount as an average value holding for an entire group of ten Ss which predicted a particular series. Only the last 120 trials are considered in this analysis.

Effect of previous events on the predictions of subjects.—We assumed that an exhaustive set of the events which preceded each prediction of each S on a particular trial included the symbols which S had predicted on previous trials and the symbols which had appeared on those trials. If only the events occurring on the preceding trial are considered, there are, then, four experiential states in which S could be at the beginning of a trial. These are (H,h), (H,v), (V,h), and (V,v), where the capital letter denotes the symbol predicted by S on the preceding trial and the small letter denotes the symbol which appeared. Each of these states has associated with it the computed probability that S will be in that state at the beginning of a trial.[2] Using this probability and the computed probability that S would predict a particular symbol when he was in each of these experiential states,[3] we computed $H_{xy}(y)$ as our average uncertainty in stating which symbol S will predict when we know what he predicted on the preceding trial and we know which symbol actually appeared. These quantities, computed separately for each group of Ss, are given in line (2) of Table 2. On line (1) the average uncertainties are presented which result when the events of the previous trial are ignored, and on line (3) the average uncertainties are presented which result when the events of the previous *two* trials are considered. Consideration of the events occurring on the preceding three trials did not further decrease our uncertainty appreciably.

Knowledge of these joint events which occurred on previous trials in our study always decreases our uncertainty in stating which symbol S is likely to predict on a given trial. This is especially true of Ss predicting Series 2 and 4 which are the two series containing sequential dependencies, but it is true also to a lesser degree of Ss predicting Series 1 and 3 which contained no sequential dependencies. Knowledge of previous events produces little reduction in our uncertainty about Ss predicting Series 1. This is not surprising as the series itself is chaotic and provides no clues as to its future behavior.

We next determined the degree to which knowledge of the separate events occurring on previous trials decreased our uncertainty. We computed $H_y(y)$ which is our average uncertainty in stating which symbol S will predict when we know *only* what he predicted on the previous trial.

[2] For example, $p(H,h)$ is computed from

$$p(H,h) = N_{H,h}/N$$

where $N_{H,h}$ is the number of trials on which S correctly predicted the H symbol and N is the total number of trials.

[3] The probability, $p_{H,h}(H)$, is computed from

$$p_{H,h}(H) = N_{H,h,H}/N_{H,h}$$

where $N_{H,h,H}$ is the number of trials on which S predicted the H symbol after correctly predicting it on the previous trial.

The behavior of the symbol series is ignored here. These quantities are given in Table 3 along with $H(y)$, our average uncertainty resulting when previous predictions are ignored, and $H_{y,\ y}(y)$, our average uncertainty resulting when the previous two predictions of *S*s are considered. With the exception of *S*s' predicting Series 1, our uncertainty in stating what *S*s will predict on a given trial is reduced in every case when we know *only* what they have predicted on previous trials. The *S*s predicting Series 3 appear to be most responsive to the predictions made on previous trials. Since this symbol series contained no sequential dependencies, we could expect that these *S*s would not find important clues about the future behavior of the series in the past behavior of the series itself. They may have looked for such clues instead in the previous predictions which they had made. This was unjustifiable since the symbol series was entirely determined in advance and independent of the predictions of *S*s. The discriminatory response of *S*s to previous responses which they had made has been reported before, however, in a situation where the memory of previous responses was the only discriminable stimulus present on each trial (**10**).

We should expect *S*s predicting Series 2 and 4, which contained sequential dependencies, to be more responsive to the past behavior of the symbol series itself. This notion was tested by computing two new quantities. These are $H_x(y)$, our uncertainty in stating what *S* will predict when we know *only* which symbol had appeared on the previous trial, and $H_{x,\ x}(y)$, our uncertainty when we know *only* which symbols had appeared on the previous two trials. The previous predictions of *S*s are ignored here. The symbols which had occurred previous to a given trial, of course, are always constant for all *S*s predicting the same series. These quantities are presented in Table 4 and approximate more closely those of Table 2. The particular symbols which had appeared on preceding trials were more important in determining predictions on succeeding trials than were the predictions which *S*s had made on preceding trials. This is especially true of *S*s predicting symbol series with sequential dependencies.

We have described three classes of events occurring on previous trials which influence the predictions which *S*s make on the next succeeding trial. These three classes of events are the predictions which *S*s made on previous trials, the symbols which appeared on previous trials, and the joint occurrence of these two events on previous trials. We have shown that our knowledge of each of these classes of events decreases our uncertainty in stating which symbol *S* will predict on

TABLE 3

THE AVERAGE UNCERTAINTIES RESULTING WHEN THE PREVIOUS PREDICTIONS OF SUBJECTS ARE KNOWN

Average Uncertainties	Series 1	Series 2	Series 3	Series 4
$H(y)$	1.00	1.00	.76	.80
$H_y(y)$	1.00	.83	.72	.75
$H_{y,\ y}(y)$	1.00	.83	.72	.73

TABLE 4

THE AVERAGE UNCERTAINTIES RESULTING WHEN THE PREVIOUS SYMBOLS, APPEARING IN THE SYMBOL SERIES, ARE KNOWN

Average Uncertainties	Series 1	Series 2	Series 3	Series 4
$H(y)$	1.00	1.00	.76	.80
$H_x(y)$	.99	.69	.74	.70
$H_{x,\ x}(y)$	.98	.55	.70	.56

TABLE 5

THE AVERAGE UNCERTAINTIES RESULTING WHEN THE CORRECTNESS OF PREVIOUS PREDICTIONS IS KNOWN

Average Uncertainties	Series 1	Series 2	Series 3	Series 4
$H(y)$	1.00	1.00	.76	.80
$H_x(y)$	1.00	.99	.74	.78
$H_{x,\ s}(y)$	1.00	.89	.70	.76

TABLE 6

THE AVERAGE UNCERTAINTIES OF THE SYMBOL SERIES

Average Uncertainties	Series 1	Series 2	Series 3	Series 4
$H(x)$	1.00	1.00	.81	.80
$H_s(x)$	1.00	.78	.80	.57
$H_{s,\ s}(x)$	.98	.77	.79	.56

the next trial. The greatest decrease in our uncertainty occurs when we know the joint occurrences of events on preceding trials (Table 2). It would be convenient and interesting if we could state the proportion of this maximum decrease in uncertainty which results when we know only *either* which symbols S predicted or which symbols appeared on preceding trials (Tables 3 and 4). We cannot do so for two reasons.

1. It is apparent from the tables that the amount by which our uncertainty is reduced by knowledge of the joint events of previous trials is not equal to the sum of the amounts by which it is reduced by knowledge of each of the two separate classes of events. It can be shown that this will never be true so long as the occur-

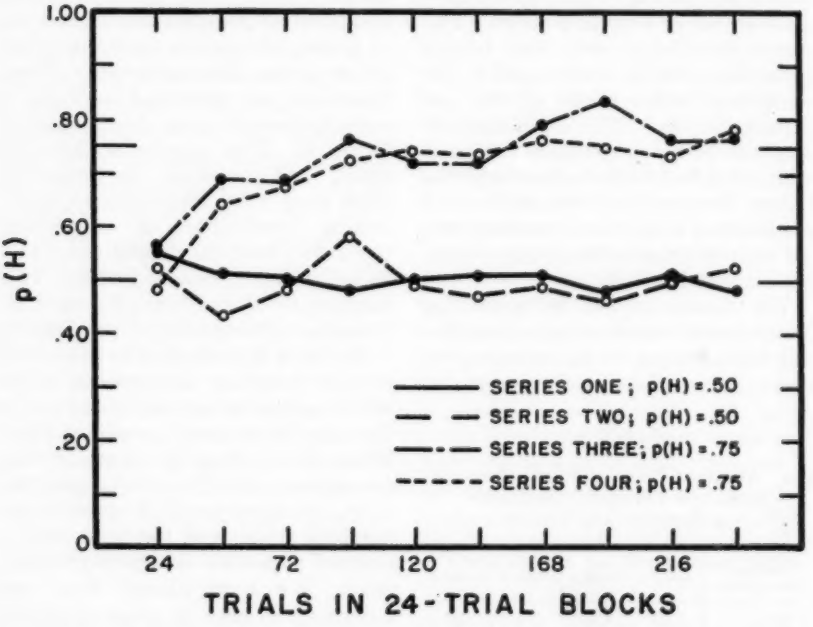

FIG. 1. The proportion of trials, in successive 24-trial blocks, on which the symbol H was predicted. The legend within the figure gives the statistical parameters of the series being predicted.

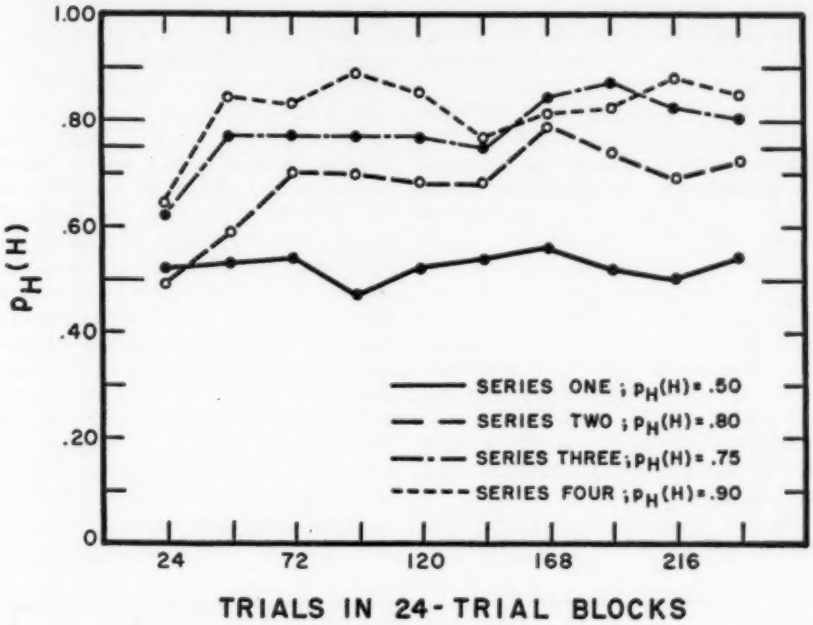

FIG. 2. The proportion of trials, in successive 24-trial blocks, on which the symbol *H* was predicted when it had been predicted on the preceding trial

rences of these separate events on previous trials are not separately and independently determined (11). We have shown that they are not independently determined since the predictions made by *S* are influenced by the symbols which have appeared (Table 4).

2. We have found that the events of previous trials which we have described so far are not exhaustive of the events which can occur on those trials, as *S* views them. In the case of three groups, our uncertainty is reduced if we know *only* whether previous symbol appearances were correctly predicted. We need not know *which* symbol was correctly anticipated. These uncertainties, $H_e(y)$ and $H_{e,\,e}(y)$ are given in Table 5.

Our average uncertainty in stating which symbol will appear in the symbol series on a given trial when we know which symbols have appeared on previous trials is given for each series in Table 6. The degree to which the symbol series was statistically structured is less in every case than were the predictions of *S*s, as shown by the uncertainties of Table 2.

Changes in prediction trends throughout the trial series.—The conclusions which we have made, of course, are based upon averaged data of the last 120 trials. Actually, each *S* did not respond in the same way every time he experienced a particular succession of events. Following the occurrence of some events the predictions of *S*s were indeterminate and showed no systematic trends. Following other events, there were striking systematic

changes in the predictions made. This was especially true of the predictions which followed trials on which a symbol had been correctly predicted.

In Fig. 1 the proportion of trials, $p(H)$, in each successive 24-trial block on which the symbol H was predicted is plotted for all prediction series. These curves are similar to those reported before for "acquisition" trials in verbal conditioning (2). These curves do not show that the sequential dependencies of Series 2 and 4 had any effect upon the predictions of our *S*s. The predictions of *S*s predicting Series 1 and 2 are identical, as are those of *S*s predicting Series 3 and 4. The predictions of the two 50% groups are significantly different from the two 75% groups ($p<.01$).

In Fig. 2 only those H predictions which follow an H prediction are plotted. In both of these figures it is apparent that the statistical characteristics of the prediction series resemble the characteristics of the symbol series.

A different result is obtained, however, when only those H predictions which follow a *correct* H prediction are plotted. In Fig. 3 we have plotted $p_{HC}(H)$, the proportion of correct H predictions which were followed by an H prediction. The predictions of the groups predicting Series 2, 3, and 4 are now almost indistinguishable. It appears that the predictions of the group watching Series 4 are somewhat more determined in the early trials, but by the halfway mark *S*s in all three groups appear to respond in about the same way once they have experienced a trial on which they had correctly anticipated the H symbol. This provides the interesting result

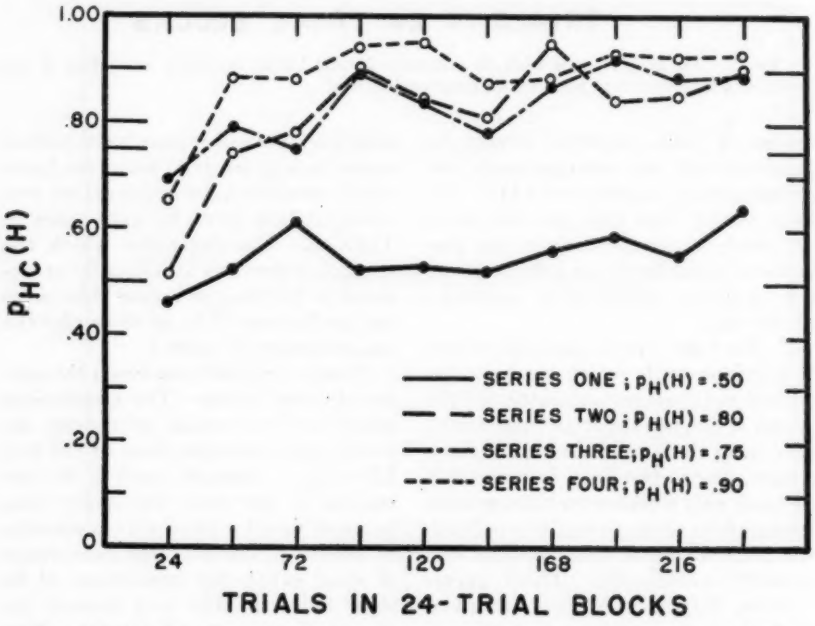

FIG. 3. The proportion of trials, in successive 24-trial blocks, on which the symbol H was predicted when it had been correctly predicted on the preceding trial

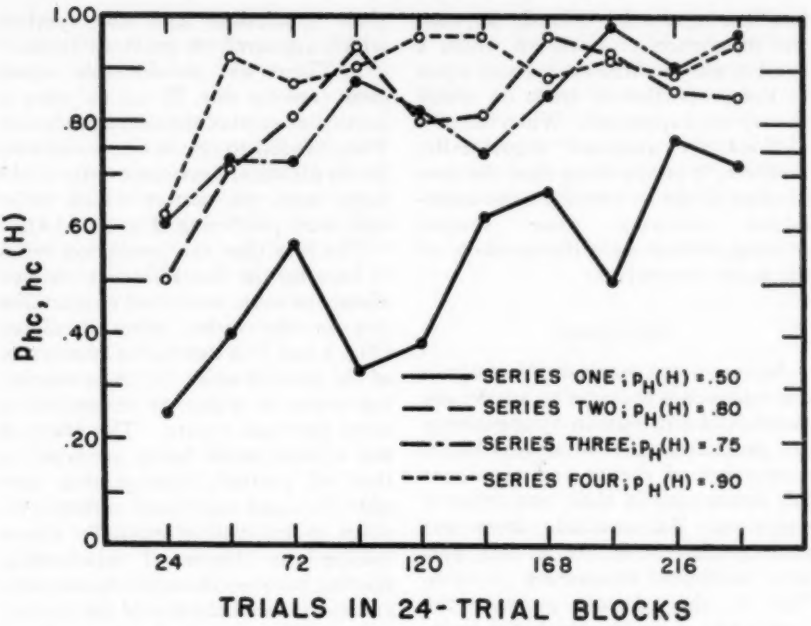

FIG. 4. The proportion of trials, in successive 24-trial blocks, on which the symbol *H* was predicted when it had been correctly predicted on the two preceding trials

that *S*s who do not predict *H* an equal number of times over a series of trials (Series 2 and 3, Fig. 1) are in agreement as to exactly *when* it should be predicted. The difference between the results of Fig. 1 and Fig. 3 can be explained, of course, by the fact that *S*s watching Series 2 and *S*s watching Series 3 do not correctly anticipate the *H* symbol with equal frequency. This frequency is determined most importantly by the statistical characteristics of the symbol series.

The transitional probability, $p_{hc,\,hc}(H)$, for having correctly predicted *H* twice in succession and then predicting *H* again is presented in Fig. 4. The *S*s predicting Series 2, 3, and 4 are still in fair agreement as to when *H* should be predicted. The *S*s watching Series 1, the chaotic series, however, now show a systematic tendency to increase the likelihood of predicting *H* following two correct anticipations of that symbol. This upward trend is statistically significant as shown by a test for trend ($\rho < .025$) (1).

Not all sequences of previous events are such adequate stimuli for succeeding predictions. There is a tendency, rather, for *S*s to respond most systematically to homogeneous runs of the same symbol, as has been reported before (5). The *S*s of every group, however, showed highly systematic response tendencies following at least some previous events. This means that the statistical characteristics of the series of predictions of *S*s were actually not in agreement with the characteristics of the symbol series they were predicting. It is true, as

has been reported previously (**2**), that the proportion of trials on which a symbol was predicted was about equal to the proportion of trials on which that symbol appeared. When the predictions are analyzed sequentially, however, it is apparent that the predictions of our *S*s were far more determined following some symbol sequences than were the symbols of the series themselves.

Discussion

An important result of this study is the evidence it provides to answer the question of how human *S*s apparently can perceive statistical structuring in a sequence of events and reproduce this structuring in their own series of responses. We conclude from our evidence that *S*s do not, in fact, perceive statistical structuring as such. That is, they do not perceive the probability rules by which the series of events was generated. They do perceive, instead, those short sequences of events which precede each prediction, which can be discriminated from other possible sequences, and which are found to provide some information about the future behavior of the symbol series. There are several interesting conclusions which we can make about the way in which *S*s perceive these previous events.

1. All combinations of possible previous events were not discriminated with equal ease. Some previous events, especially homogeneous runs of the same symbol, were more easily discriminated and consistently responded to than were others.

2. The previous events to which our *S*s responded on each trial included more than just the symbols which had been appearing. They included also the previous predictions of *S*s and the degree of correspondence between their predictions and the symbols which appeared on previous trials.

3. There was considerable agreement among our *S*s as to *when* a particular symbol should be predicted. They tended to respond to some similar or identical previous events in the same way, no matter which series they were predicting (Fig. 3 and 4).

The fact that the prediction series of *S*s, after the first 120 trials, showed about the same statistical characteristics as the series being predicted (Fig. 1 and 2) is due to the interaction of the symbol series and these consistent ways in which *S*s responded to some previous events. The effect of the symbol series being predicted is that of partially determining how *often* *S*s could experience certain previous events and of partially determining the degree of relationship existing between these previous events and the future behavior of the symbol series. The statistical characteristics of the series of predictions of *S*s are then at least partially determined by these two variables: (*a*) the number of times that *S* is given an opportunity to respond in a consistent way to discriminable sequences of previous events, and (*b*) the degree of success *S* has in predicting future symbol behavior by responding consistently to discriminable sequences of previous events. The statistical characteristics of the response series of *S*s are, then, the *result* of these stimulus-response relationships and not the cause of them. That is, *S*s do not *first* perceive the rules by which the series were constructed and then respond accordingly.

We believe that this notion is most in accord with our data and accounts for two factors which have not been explained previously. First, it has been noted that *S*s do not act to maximize correct predictions by con-

sistently predicting the symbol whose probability of occurrence is greater than .50 (2). Second, it has been reported that *Ss* can make a particular response a given proportion of the time without counting their previous responses and without being aware that they are responding according to any statistical rule (10). The explanation for both of these results lies in the fact that *Ss* do not act like maximizing functions or like statistically governed mechanisms at all. The statistical characteristics of their responses are only by-products of the stimulus-response relations we have suggested.

The results of this study and this explanatory notion have important implications for the concept of information which is derived from information theory. It has been stated that the informational value of the occurrence of one event from a set of events is a function of the likelihood of occurrence of that event (11). The likelihood of occurrence here is defined by the statistical rules by which this event is chosen from the set. This concept of information has had important applications where only physical systems are involved which select symbol after symbol, according to statistical laws, for transmission to a receiving system. When a human observer becomes the receiver of such a message, composed of a series of symbols chosen in that way, however, we can no longer specify the informational value of the message to the observer by stating the statistical laws by which it was generated *unless the degree to which the observer expects each symbol before it occurs conforms exactly to these statistical laws*. We have shown, essentially, that this condition does not hold. It does not hold because the observer does not directly perceive the rules by which a series of events is generated. That is, he does not perceive an apparently random series as being homogeneously random, but responds instead as though the series were composed of small subsequences, some of which are dependable cues to the future behavior of the series.

Furthermore, the observer does not predict the future behavior of a series of events purely on the basis of the past behavior of the series. The observer contaminates his judgment by including his own previous behavior and the "correctness" of his own previous behavior in the set of events which influences his expectancy of future events.

This last point is important, we believe, because it means, if our *Ss* are typical, that *Ss* will always perceive an ambiguous series of events as being more structured than it really is. This must follow whenever *Ss* allow their own past behavior to influence their predictions of the future occurrence of series of events which are, in fact, completely independent of the behavior of *Ss*.

Summary

The *Ss* were required to predict on each of 240 trials which symbol of two possible symbols, *H* or *V*, would appear on that trial. Each of four groups of *Ss* predicted a different series of symbols. The four symbol series differed from each other in terms of the proportion of trials on which a particular symbol appeared and in the degree of sequential dependency existing between successive symbols.

The results were as follows:

1. The *Ss* in all groups began predicting the two symbols about equally often and gradually adjusted their predictions to conform to the actual prob-

ability of occurrence of the symbols in the symbol series. The *S*s tended also to predict a symbol, following a previous prediction of that symbol, about as often as the symbol succeeded itself in the symbol series.

2. The predictions of *S*s were sequentially analyzed to determine the degree to which their predictions on a given trial were determined by the events occurring on preceding trials. The analysis showed that the specific stimuli to which *S*s responded on each trial were the sequences of events, including the previous predictions of *S*s and the correctness of these predictions, which had occurred on the previous two trials. Some sequences of previous events, especially homogeneous series of a correctly predicted symbol, were shown to be more adequate stimuli than were others.

3. The apparent ability of *S*s to perceive and reproduce sequential statistical structuring was explained in terms of response to specific sequences of events occurring prior to each trial. The relevance of these results to the formulation of information transfer was discussed.

(Received for early publication
September 29, 1952)

References

1. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.
3. Grant, D. A., Hornseth, J. P., & Hake, H. W. The influence of the inter-trial interval on the Humphreys' "random reinforcement" effect during the extinction of a verbal response. *J. exp. Psychol.*, 1950, **40**, 609–612.
4. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
5. Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.
6. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
7. Lewis, D. J. Partial reinforcement in a gambling situation. *J. exp. Psychol.*, 1952, **43**, 447–450.
8. Miller, G. A., & Frick, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.
9. Newman, E. B. Computational methods useful in analyzing series of binary data. *Amer. J. Psychol.*, 1951, **64**, 252–262.
10. Senders, U. L., & Sowards, A. Analysis of response sequences in the setting of a psychophysical experiment. *Amer. J. Psychol.*, 1952, **65**, 358–374.
11. Shannon, C. E. A mathematical theory of communication. *Bell Syst. tech. J.*, 1948, **27**, 379–423.

EARLY PUBLICATION

IN

APA JOURNALS

The policy of accepting articles for immediate publication (providing the editor accepts the article and the author is willing to pay the entire cost of increasing the next available issue by enough pages to add his article to the normal content) is now standard practice for all APA journals except *Psychological Abstracts* and the *American Psychologist.*

The actual charge made to the author includes three items:

1. A basic charge of so much per page. This is the minimum amount that it costs to add an additional page to the journal. For 1951 these costs are:

	Per Page
Journal of Abnormal and Social Psychology	$17.00
Journal of Applied Psychology	15.00
Journal of Comparative and Physiological Psychology	11.00
Journal of Consulting Psychology	14.00
Journal of Experimental Psychology	14.00
Psychological Bulletin	17.00
Psychological Monographs: General and Applied	15.00*
Psychological Review	14.00

* Since each Psychological Monograph is printed separately, the author of one handled on an early publication basis can be charged exactly the cost of printing. The figure of $15.00 is an approximate one; the actual figure will be higher for very short monographs and lower for very long ones. The cost will also vary depending upon the amount of special composition and the illustrations used.

These charges are based upon several factors:

(a) The greater number of words on a particular journal page, the higher the cost per page. Conversely, the fewer words printed on the page, the lower the cost per page.

(b) The more copies which must be printed, the higher the cost.

(c) The more expensive the printer, the higher the cost. Compared to the factors listed above, this is not an important difference in the charges made.

2. The full cost of any cuts or other illustrative material, of special composition for tables, and of author's changes in proof.

3. The full cost of any reprints ordered. (Authors of early-publication articles do not receive any free reprints.)